



ERICKSON LEON KOVALSKI

DIÁLOGOS ENTRE WETWARE E INTELIGÊNCIA ARTIFICIAL
Implicações das concepções sobre processos cognitivos biológicos no
desenvolvimento da tecnologia computacional.

Dissertação apresentada ao Programa de Pós-Graduação em História das Ciências e das Técnicas e Epistemologia – HCTE, Universidade Federal do Rio de Janeiro – UFRJ como parte dos requisitos da obtenção do grau de Mestre.

Orientador: Prof. Dr. Luiz Pinguelli Rosa

Rio de Janeiro
2021

CIP - Catalogação na Publicação

KK88d Kovalski, Erickson Leon
Diálogos entre o Wetware e Inteligência Artificial: Implicações das concepções sobre processos cognitivos biológicos no desenvolvimento da tecnologia computacional / Erickson Leon Kovalski. -- Rio de Janeiro, 2021.
125f.

Orientador: Luiz Pinguelli Rosa.
Dissertação (mestrado) - Universidade Federal do Rio de Janeiro, Decania do Centro de Ciências Matemáticas e da Natureza, Programa de Pós-Graduação em História das Ciências e das Técnicas e Epistemologia, 2021.

1. Inteligência Artificial. 2. História da Computação. 3. Teoria da Mente. 4. Aprendizagem. 5. Inteligência. I. Pinguelli Rosa, Luiz , orient. II. Título.

DIÁLOGOS ENTRE O WETWARE E INTELIGÊNCIA ARTIFICIAL:
IMPLICAÇÕES DAS CONCEPÇÕES SOBRE PROCESSOS COGNITIVOS BIOLÓGICOS NO
DESENVOLVIMENTO DA TECNOLOGIA COMPUTACIONAL

Erickson Leon Kovalski

DISSERTAÇÃO APRESENTADA AO CORPO DOCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM HISTÓRIA DAS CIÊNCIAS E DAS TÉCNICAS E EPISTEMOLOGIA – HCTE, DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO – UFRJ, COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM HISTÓRIA DAS CIÊNCIAS.

BANCA EXAMINADORA

Orientador: Prof. Dr. Luiz Pinguelli Rosa, UFRJ

Prof. Dr. Rodrigo Siqueira Batista, UFV e FADIP

Prof^a. Dr^a. Virgínia Maria Fontes Goncalves Chaitin, UFRJ

Prof. Dr. Nelson Job Vasconcelos de Carvalho, UFRJ

RIO DE JANEIRO, RJ - BRASIL
10 DE MARÇO DE 2021

AGRADECIMENTOS

Agradeço a todos que de algum modo auxiliaram no desenvolvimento desta pesquisa, com contribuições materiais e intelectuais.

*A few years ago, Marvin Minsky made a remark that many people, especially those who believe that thought—intelligence—can never be mechanized, found deeply troubling. Minsky said, “**The brain happens to be a meat machine.**” Minsky has a fairly low tolerance for what he considers to be nonsense. The idea that it is somehow impossible to understand the mind he regards as nonsense.*

— Jeremy Bernstein

RESUMO

Em uma abordagem histórica, o pensamento mecânico, que levou à criação do ramo da Ciência da Computação chamado Inteligência Artificial, tem sua origem nos mitos da Grécia homérica e começa a ser desenvolvido através de diversos avanços na lógica. Na modernidade Leibniz infunde ideais de uma língua da ciência e do cálculo como meio de resolver divergências na ciência. Máquinas que implementam o ideal de Leibniz são os precursores do computador digital contemporâneo. A aprendizagem pode essencialmente ser descrita mecanicamente de forma que compreenda desde as formas mais básicas, como a aquisição de conhecimento do mundo na evolução biológica às formas artificiais na forma de redes neurais. O conceito de inteligência é complexo e difícil de definir, porém incorpora adaptabilidade ao ambiente e flexibilidade para tratar de situações novas. Para abranger humanos, animais e máquinas a noção de inteligência como capacidade de prever o ambiente parece ser mais produtiva. A inteligência é dependente do problema, e no contexto de cada problema se apresenta como um algoritmo de otimização.

Palavras-chave: Inteligência Artificial, Teoria da Mente, Lógica, Cognição, Evolução.

ABSTRACT

In a historical approach, the idea of mechanical thinking led to the creation of the branch of Computer Science called Artificial Intelligence. It originated in the myths of Homeric Greece and had been developed through several advances in the field of logic. In modernity, Leibniz instilled the idea of a language of science and calculation as a means of resolving disputes in science. Machines that carry out Leibniz's ideal are the precursors of the contemporary digital computer. Learning can essentially be described mechanically in ways that range from the most basic forms, such as acquiring knowledge of the world in biological evolution to artificial forms like neural network training. The concept of intelligence is complex and difficult to define, but it incorporates adaptability to the environment and flexibility to deal with novel situations. To encompass humans, other animals and machines, the notion of intelligence as the ability to predict the environment seems to be more productive. Intelligence is dependent on the problem, and on the context of each problem, and can be described as an optimization algorithm.

Key words: Artificial Intelligence, Theory of mind, Logic, Cognition, Evolution.

LISTA DE FIGURAS

FIGURA 1.....	24
FIGURA 2.....	32
FIGURA 3.....	34
FIGURA 4.....	37
FIGURA 5.....	39
FIGURA 6.....	44
FIGURA 7.....	45
FIGURA 8.....	73
FIGURA 9.....	81
FIGURA 10.....	94

LISTA DE QUADROS

QUADRO 1.....	26
QUADRO 2.....	27
QUADRO 3.....	27
QUADRO 4.....	34
QUADRO 5.....	34
QUADRO 6.....	35
QUADRO 7.....	36
QUADRO 8.....	55
QUADRO 9.....	63
QUADRO 10.....	79

LISTA DE TABELAS

TABELA 1:	42
TABELA 2	43

LISTA DE ABREVIATURAS E SIGLAS

Abreviação	Significado	Tradução para o Português (se terminologia estrangeira)
ACO	Ant Colony Optimization	Optimização baseada em colônia de formigas
AEC	Antes da Era Comum	
AGI	Artificial General Intelligence	Inteligência Artificial Geral
AI	Artificial Intelligence	Inteligência Artificial
EC	Era Comum	
FOL	First Order Logic	Lógica de Primeira Ordem
g, g_f, g_c	Factor g , fluid intelligence factor of g , crystallized intelligence factor of g	Fator g , fator de inteligência fluida em g , fator de inteligência cristalizada em g
IQ	Intelligence Quotient	Quociente de Inteligência
ML	Machine Learning	Aprendizado de máquinas
PAC	Probably Approximately Correct	Provavelmente aproximadamente correto
PM	Principia Mathematica	
SOL	Second Order Logic	Lógica de Segunda Ordem
TSP	Traveler's Salesman Problem	Problema do caixeiro viajante

SUMÁRIO

APRESENTAÇÃO	13
1 INTRODUÇÃO	15
2 EM BUSCA DO PENSAMENTO MECÂNICO	19
2.1 Introdução	19
2.2 Mecanismos da Antiguidade e Medievo	20
2.3 Mecanismos da Modernidade.....	21
2.4 Pensamento mecânico na Antiguidade e Medievo	25
2.5 Pensamento mecânico na Modernidade	28
2.6 Lógica.....	33
2.7 Pioneiros.....	46
2.8 Conclusão e Discussão.....	47
3 APRENDIZAGEM EM MÁQUINAS: ORGANISMOS BIOLÓGICOS E INTELIGÊNCIA ARTIFICIAL	49
3.1 Introdução	49
3.2 A Inteligibilidade do Real.....	51
1.1.1. UM CAMINHO PARA SE ENTENDER APRENDIZAGEM.....	53
3.3 Evolução Biológica	54
3.3.1 SIMULAÇÃO DA EVOLUÇÃO BIOLÓGICA	58
3.3.2 COMPUTAÇÃO EVOLUTIVA	61
3.3.3 MODELO PROVAVELMENTE APROXIMADAMENTE CORRETO.....	63
3.4 Conclusão e Discussão.....	66
4 INTELIGÊNCIA EM MÁQUINAS: PODEM MÁQUINAS PENSAR?	68
4.1 Introdução	68
4.2 Algumas dificuldades em se estudar a inteligência.....	69
4.3 A Inteligência Humana	75
4.3.1 DEFINIÇÕES DE INTELIGÊNCIA HUMANA.....	76
4.3.2 MENSURAÇÃO DA INTELIGÊNCIA HUMANA.....	77
4.4 Inteligência em Inteligência Artificial	82
4.5 Conclusão	89
4.6 Discussão	91
5 CONCLUSÃO	96
REFERÊNCIAS	97
APÊNDICES.....	106

APRESENTAÇÃO

Serão apresentados aqui, em seguida, três textos com diferentes perspectivas sobre um tema comum, que é a Inteligência Artificial (AI). O formato escolhido permite que cada um dos textos possa ser futuramente descolado da dissertação e transformado em um artigo para que seja submetido à publicação em um periódico científico. Algumas repetições e reapresentações de ideias acontecem em decorrência desta proposta. Também, em muitos casos, apresentamos certas ideias que, em um artigo, seriam consideradas pré-requisitos para leitura. Isto aconteceu, principalmente, pela riqueza do material coletado durante o mestrado, criando, assim, pontos de inserção para a continuidade desta pesquisa no futuro.

No primeiro texto, intitulado *Em busca do pensamento mecânico: uma breve história das origens da Inteligência Artificial*, foi proposta uma abordagem histórica, procurando alguns elementos que reconstroem o nascimento da AI como campo de estudo. A forma do texto foi inspirada pela obra de Pamela McCorduck, *Machines Who Think: A Personal Inquiry Into the History and Prospects of Artificial Intelligence*, de 1979. Esta obra, como o título revela, tem um caráter de investigação pessoal, através deste olhar, os diversos eventos e personalidades históricas são alinhavados. A segunda obra que nos serviu de inspiração foi o livro em três volumes *Tecnociências e humanidades: novos paradigmas, velhas questões* de nosso orientador, professor Dr. Luiz Pinguelli Rosa, que usando uma cuidadosa revisão de referências construiu um caminho histórico desde a antiguidade à física contemporânea. Grande parte das questões trabalhadas emergiram da leitura e discussão desta obra, assim como a intenção de fazer do texto uma lista de referências para futuro aprofundamento nos temas abordados.

No segundo texto, *Aprendizagem em máquinas: organismos biológicos e Inteligência Artificial*, procura-se entender o mecanismo de aprendizagem, em máquinas e na biologia. O texto é mais especulativo e com algumas preocupações epistemológicas, como o limite do conhecimento e o problema de indução. Um interesse, nessa abordagem, foi cruzar a visão de profissionais da Biologia, quando falam de computação, e da Ciência da Computação, quando se voltam para o mundo

biológico. Talvez a melhor maneira de descrever este trabalho seja caracterizando-o como uma justaposição, dialogada, de ideias transdisciplinares.

O último texto, *Inteligência em máquinas: podem máquinas pensar?* discorre sobre o conceito de inteligência de uma forma geral, e sobre a possibilidade de uma definição abrangente o suficiente para incluir o conceito da Psicologia humana, da Cognição Animal e da Ciência da Computação. Sendo a inteligência o tema central que motivou esta pesquisa de mestrado, o intuito foi tentar capturar o melhor que pudéssemos este conceito. As preocupações deste texto tiveram origem nas discussões durante a disciplina de *Cognição e Computação* (I e II) do professor Sérgio Exel Gonçalves, no programa de *Pós-graduação em História das Ciências, Técnicas e Epistemologia* da UFRJ, ministradas no ano de 2017.

As siglas foram mantidas, na maior parte, no formato em inglês por se tratar de jargões da área. A tentativa de introduzir siglas da tradução dos termos para o português poderia causar desnecessária dificuldade ao leitor já acostumado com textos da área. Todavia, uma lista de siglas foi incluída para referência, com indicações em português de seu significado.

Quanto às citações longas, procuramos padronizar usando sempre a obra em inglês. Quando o idioma original não era o inglês, procuramos uma tradução para este idioma. As citações no corpo do texto foram traduzidas da obra consultada. A maioria dos livros citados foram usados em seu formato digital, portanto não possuem número de páginas. Para encontrar a citação em contexto é, portanto, necessário usar a ferramenta de pesquisa – e nisto o fato de estarem em língua inglesa facilita bastante. No entanto, há indicação de páginas nos casos em que obras físicas, ou digitalizações de obras físicas, foram utilizadas.

Ilustrações, quando não indicadas como *creative commons* ou domínio público, foram utilizadas minimamente e sob a expectativa de *fair use* em se tratando de trabalho acadêmico sem finalidade comercial. A atribuição dos autores foi indicada em todos os casos. Apenas as ilustrações de nossa elaboração não contêm indicação autoral.

1 INTRODUÇÃO

The chessboard is the world; the pieces are the phenomena of the universe; the rules of the game are what we call the laws of Nature. The player on the other side is hidden from us. We know that his play is always fair, and patient. But also we know, to our cost, that he never overlooks a mistake, or makes the smallest allowance for ignorance.

Thomas Henry Huxley

Organismos biológicos são capazes de resolver um inumerável rol de problemas com os quais se deparam e de cuja solução depende sua sobrevivência. Muitos destes problemas fazem parte de uma classe que requer uma abordagem matemática bastante difícil, chamados problemas NP-hard. Isto é, problemas nos quais a complexidade cresce exponencialmente em relação ao aumento do número de elementos envolvidos. Esta é uma dificuldade que aparece frequentemente no desenvolvimento de tecnologias de Inteligência Artificial (AI). Uma forma de exemplificar este tipo de problema é imaginar uma situação em que seja necessário calcular o menor trajeto para visitar várias cidades e voltar ao lugar de partida. Se pensarmos em um pequeno número de trajetos, como sete cidades, podemos resolver, calculando todos os trajetos possíveis, que somam 360 diferentes possibilidades e, dentre estes, escolher o menor. Porém se fossem 8 as cidades, o número de trajetos possíveis seria de 2520. Logo percebe-se que a técnica da força bruta será ineficiente para um número muito elevado de cidades. Este é conhecido como *Traveling Salesman Problem* (TSP), ou Problema do Caixeiro Viajante, um exemplar da classe NP-Hard¹. Uma forma de abordar estes problemas é imitar as soluções naturais, pois organismos biológicos resolvem problemas similares e de muito alta complexidade o tempo todo. Estas formas de otimização natural podem ser modeladas computacionalmente na forma de um algoritmo. No caso, uma abordagem possível (dentre muitas) é o *Ant Colony Optimization* (ACO), ou

¹A importância do tema P vs NP levou à produção de um artigo, desenvolvido durante um curso de pós-graduação em Tecnologia da Informação, independente do mestrado, porém guardando fundamental relação com a temática desta Dissertação. Em razão disso, parte do trabalho foi incluído na dissertação como apêndice (APÊNDICE A) e sugerido como auxiliar para o entendimento de alguns tópicos.

otimização baseada no comportamento de formigas procurando alimento (APÊNDICE C).

Complexidade, aqui, entenderemos como algo ligado ao tamanho de uma mensagem, descontadas as suas redundâncias. A partir de seu artigo seminal de 1948, Shannon elabora um método de codificação que utiliza propriedades estatísticas dos símbolos da mensagem para construir uma versão reduzida da mensagem, que porém possui toda a informação original (SHANNON, 1948). Kolmogorov em 1965 utiliza da teoria da computação para quantificar a informação (KOLMOGOROV, 1968). Assim, a quantidade de informação de uma mensagem seria o tamanho do menor programa para produzi-la. Não obstante, de acordo com os resultados de Turing (TURING, 1936) seria impossível algoritmicamente se descobrir genericamente qual é este menor programa, mas é possível gradualmente se aproximar deste resultado.

A matemática frequentemente depara-se com problemas equivalentes, isto é, a solução para um impasse em um problema pode vir da demonstração que este problema pode ser formulado de outra forma, a qual tem solução (em positivo ou negativo) ou ao menos é mais fácil de se abordar. De certa forma, é como dizer que o mesmo tipo de problema aparece em diversas formulações, porém pode ocorrer que uma representação é mais frutífera que outra.

Supomos, de forma otimista, que os problemas corriqueiros que organismos biológicos enfrentam tenham entre si certa relação. Problemas os quais incidentalmente se mostraram os grandes desafios da AI – seja o processamento visual, o da linguagem natural, o problema da organização do conhecimento, entre inúmeros outros. Há indícios de que certas partes especializadas do cérebro podem ser cooptadas para desempenhar diferentes funções em surdos (FINNEY, FINE, *et al.*, 2001), que partes especializadas em uma modalidade sensorial podem aprender a responder por dados de outra modalidade (ROE, PALLAS, *et al.*, 1992), e que há similaridades estruturais entre partes do córtex (MOUNTCASTLE, EDELMAN, 1978). Não é um grande passo levantar a hipótese de que há algo como um algoritmo básico implementado pelo sistema nervoso (HAWKINS, BLAKESLEE, 2004).

Para este fim, a AI, de uma forma geral, pode não apenas potencializar nossa capacidade de interagir com o mundo, mas também ajudar a entender muitos comportamentos animais e mesmo desvendar o funcionamento da inteligência humana. Assim como organismos buscam otimizar o uso de recursos, a AI é uma ferramenta capaz de encontrar soluções ótimas (ou suficientemente próximas do ótimo) em um ambiente complexo e com restrições de poder e tempo computacionais (DORIGO, MANIEZZO, *et al.*, 1996, DORIGO, STÜTZLE, 2004).

Esta hipótese, levada ao extremo, é representada pela atual discussão, trazida à frente por Pedro Domingos, de que existiria um algoritmo de AI capaz de propiciar soluções satisfatórias a todos os problemas que podem ser abordados computacionalmente, que ele denominou o algoritmo mestre (DOMINGOS, 2015).

Por outro lado, elaborar modelos de soluções de problemas com base no comportamento dos organismos biológicos pode impulsionar avanços no desenvolvimento de tecnologias em AI, sobretudo frente às limitações encontradas nesta área de pesquisa, como por exemplo, a pesquisa em Inteligência Artificial Geral (AGI). Percebe-se que organismos biológicos frequentemente servem de modelo e estes têm, historicamente, guiado pesquisadores em AI, assim como em áreas correlatas², na elaboração de suas teorias, apesar de muito pouco desta influência apareça claramente no desenvolvimento e utilização das diversas tecnologias.

As técnicas utilizadas em AI, quando desenvolvidas para um problema específico, frequentemente acabam restritas a este único tipo de problema. Uma prática estabelecida é utilizar-se de um conjunto de ferramentas especializadas, sob o controle de um gerenciador, o qual decide qual destas tem a melhor solução por meio de um cálculo de probabilidades. O ponto fraco desta abordagem é que implica uma costura de técnicas que guardam pouca ou nenhuma similaridade entre si, acarretando a impossibilidade de se transferir o conhecimento e expertise adquiridos entre os componentes do sistema.

² Em sentido amplo, AI inclui desde a perspectiva simbolista às redes neurais, passando por abordagens como a dos sistemas especialistas e ontologias. Porém, várias razões levaram a certas abordagens não serem mais entendidas como parte da AI. Existe certo efeito de que problemas já resolvidos imediatamente deixam de ser parte da AI. Atualmente existe a tendência de se entender como AI apenas as técnicas de machine learning. Não será esta a abordagem deste trabalho, no qual se pretende permanecer mais próximo ao ideal dos fundadores.

A evolução biológica, no entanto, utiliza-se das estruturas já presentes para construir novas habilidades. O tempo todo, estruturas arcaicas estão sendo reaproveitadas através de uma série de pequenas adaptações a novas situações do ambiente, conferindo maior aptidão para sobrevivência e, conseqüentemente, reprodução do organismo – que se denomina *fitness*. Em outras palavras, as expertises dos sistemas são parcialmente transferidas para novas aplicações, diferentemente do observado na AI.

O que, superficialmente, parece ser o caso de uma dificuldade técnica, em um nível mais profundo se conecta aos grandes problemas da História da Filosofia, ao se questionar como as operações realizadas por dispositivos eletrônicos poderiam adquirir conteúdo semântico, produzir subjetividade e um *sensu de si*. Como a análise de faixas espectrais de luz em um sensor poderiam ser entendidas como sensação de cor? Como operações mecânicas – que podem ser descritas através de operações simples sobre números – adquiririam a riqueza de experiências típicas dos seres biológicos? Máquinas movidas à vapor, relés, transistores, ou um sistema de gatos, ratos e pedaços de queijo (SEARLE, 1990) podem, para todos os efeitos, implementar um sistema capaz dos mesmos atributos que cérebros – com ou sem seus devidos corpos – implementam? Um dos temas centrais da Filosofia da Mente, conhecido como “problema mente corpo”, tem na Ciência da Computação uma abordagem capaz de encontrar uma resolução satisfatória no âmbito da Filosofia? É possível uma prova matemática disto? Por fim, como o fenômeno da inteligência e consciência situa-se em na procura de uma possível “Teoria de Tudo”?

2 EM BUSCA DO PENSAMENTO MECÂNICO: UMA BREVE HISTÓRIA DAS ORIGENS DA INTELIGÊNCIA ARTIFICIAL

And yet what do I see from the window but hats and coats which may cover automatic machines? Yet I judge these to be men.

Descartes

2.1 Introdução

Procuramos neste trabalho concatenar alguns elos que conectam a mitologia e o nascimento do pensamento racional na Grécia antiga ao escolasticismo e misticismo medievais, ao lado das máquinas mecânicas, analógicas, de calcular da antiguidade, medievo, às máquinas mecânicas, digitais e eletrônicas dos séculos XIX e XX.

Acreditamos que a lógica seja um desses elos, outros – não de todo disjuntos – são uma visão mecânica do mundo e certo fascínio pelos próprios processos da razão. E, no meio dessa cadeia, encontramos o filósofo e polímata Leibniz, e suas importantes contribuições tanto para a filosofia e como para a matemática. O pensador racionalista foi educado em um mundo ainda fortemente marcado pelo aristotelismo, sobretudo no viés escolástico. Leibniz, não obstante, propôs ideias inovadoras que influenciariam o mundo por séculos e, algumas destas ideias, como pretendemos apontar, desembocaram na criação da Ciência da Computação.

Um primeiro tema que nos interessa é a ideia de que objetos inanimados poderiam ser, de alguma forma, construídos para desempenhar funções que, com base em um conceito exclusivista de inteligência, se acreditava apenas humanos – ou talvez seres supra-humanos – seriam capazes. A arte a serviço deste propósito poderia estar nos poderes divinos, no misticismo ou bruxaria, poderia estar na elevada – e incompreensível para alguns – erudição de alguns acadêmicos, ou, mais recentemente, na hermeticidade da Matemática, Estatística e Ciência da Computação. De fato, Harold Abelson e Gerald Jay Sussman, no influente livro *Structure and Interpretation of Computer Programs*, comparam as sentenças do LISP – a linguagem de escolha dos primeiros pesquisadores da Inteligência Artificial – ao ato de proferir encantamentos. E os métodos atuais em redes neurais são tão complexos e não documentados que muitos descrevem como *magia negra*.

Não obstante, nada há de hermético ou incompreensível – ao menos é isto que pretendemos mostrar aqui – com um resgate da genealogia da Inteligência Artificial na lógica, nas máquinas de calcular e na visão mecanicista de mundo. A lógica como a mecânica da razão e as máquinas como a materialização da lógica.

2.2 *Mecanismos da Antiguidade e Medievo*

A ideia de que há algo de similar entre as máquinas criadas através da engenhosidade humana e os organismos biológicos perpassa a história desde ao menos a antiguidade clássica. Ao deus Hefesto (Séc. III AEC) é atribuída a criação de Talos, um autômato metálico gigante que protegia a ilha de Creta dos invasores. A Hefesto são atribuídos, ainda, outros inventos em metal – que geralmente é o bronze –, como servos trípodes, servos com rodas e mesmo duas servas femininas com força física, razão e fala, para lhe ajudar na oficina (RHODIUS, 2009, TRUITT, 2015).

Na tradição judaica da Idade Média aparecem as narrativas do Golem, uma figura antropomórfica produzida com barro ou lama e posteriormente animada. Estas remontam ao texto talmúdico da criação de Adão, passando pela descrição de uma serva doméstica, supostamente criada por Ibn Gabirol (Séc. XI), e pela lenda do Golem, protetor de Praga, que haveria sido construído por Rabbi Juda ben Loew (Séc XVI)(DAN BILEFSKY, 2009).

Lendas de cabeças de metal falantes aparecem em diversas instâncias durante a Idade Média, frequentemente envoltas por uma atmosfera de alquimia, mágica e poderes demoníacos. Truitt (2015) comenta que autômatos durante o medievo eram exóticos à cultura latina, chegando à Europa como presentes de autoridades das culturas islâmicas e como curiosidades trazidas por viajantes, despertando fantasias e medo no mundo cristão.

A Albertus Magnus (1200-1280) é atribuída uma história de que dedicou trinta anos de sua vida na construção de uma cabeça falante – ou em algumas versões um androide completo –, em bronze, que saudaria os transeuntes e falaria

sem parar³. E nas mesmas fontes atribui-se a destruição da máquina pelo seu discípulo Thomas de Aquinas (1255-1274) após a morte do mestre, talvez, por falar demais (TRUITT, 2015). McCorduck (2004) nota, porém, que Albertus Magnus sobreviveu a Thomas de Aquinas por seis anos. Wood (2002) comenta, em seu livro *Edison's Eve*, que da mesma forma que gatos eram associados às bruxas, cabeças de bronze falantes se associaram a homens letrados, realçando a ambivalência entre fascínio e temor.

2.3 Mecanismos da Modernidade

Foi apenas após muitos séculos que a engenhosidade mecânica deixou de ser vista como afim à bruxaria e se tornou o modelo para o entendimento do mundo encabeçado por René Descartes (1596–1650). Proclamado fundador do dualismo moderno, é também responsável por dar origem a uma forma de monismo, segundo a qual a natureza se comporta exclusivamente de forma mecânica, determinada por causas tal qual o funcionamento de um relógio. A invenção do relógio mecânico, principalmente com o uso da mola espiral, fez com que a complexidade dos movimentos automatizáveis pudesse ser aumentada quase indefinidamente – o relógio astronômico de Praga é um exemplo. Em seu texto *Meditações*, de 1641, Descartes comenta que havia observado, de sua janela, um homem e imaginado que sob seu chapéu e casaco houvesse apenas uma máquina movida por molas (DESCARTES, 1955).

Uma visão similar foi defendida por Thomas Hobbes (1588–1679) e estendida para abranger também o funcionamento da mente. Sua perspectiva, apresentada no texto *De Corpore*, de 1655, advoga que toda a realidade é material e o ato de raciocinar se dá através de computações definidas por operações aritméticas sobre símbolos (HOBBS, 1839). A razão seria, portanto, nada mais que o ato de contabilizar as consequências, através de uma série de somas e subtrações de nomes, os quais foram convencionados para significar nossos pensamentos (DYSON, 2012).

³ Há muitas versões, as referências são obscuras e difíceis de localizar, algumas vezes atribuindo-se este invento a Roger Bacon ou a ambos. Outras histórias similares que encontramos envolvem Robert Grosseteste, Arnaldus de Villanova, Don Henrique de Villena e o papa Silvestre II.

A Descartes, similarmente, também é atribuída a posse de um androide ou, mais precisamente, um ginoide⁴. Das várias versões desta lenda, reproduzimos tal como foi exposta por Gaukroger:

Since the eighteenth century, there has been in circulation a curious story about Descartes. It is said that in later life he was always accompanied in his travels by a mechanical life-sized female doll which, we are told by one source, he himself had constructed 'to show that animals are only machines and have no souls'. He had named the doll after his illegitimate daughter, Francine, and some versions of events have it that she was so lifelike that the two were indistinguishable. Descartes and the doll were evidently inseparable, and he is said to have slept with her encased in a trunk at his side. Once, during a crossing over the Holland Sea some time in the early 1640s, while Descartes was sleeping, the captain of the ship, suspicious about the contents of the trunk, stole into the cabin and opened it. To his horror, he discovered the mechanical monstrosity, dragged her from the trunk and across the decks, and finally managed to throw her into the water. We are not told whether she put up a struggle. (GAULKROGER, 1995)

Jacques de Vaucanson (1709-1782) adquiriu imenso prestígio através da construção de máquinas que imitavam animais e pessoas, sendo duas de suas obras mais memoráveis: um pato, movido por corda, que era capaz de andar, comer e defecar, supostamente digerindo o alimento; e um flautista, movido por vapor, capaz de imitar realisticamente um ser humano tocando o instrumento (WOOD, 2002).

Descartes encontrou um grande obstáculo para conectar o mundo físico aos fenômenos do pensamento e, para isso, usou de várias hipóteses concernentes à glândula pineal que, mesmo em sua época, já haviam sido mostradas equivocadas por Niccolò Massa (1485–1569), e Andreas Vesalius (1514 – 1564), e rejeitadas até por Galeno (Séc. II EC). Não conhecendo o sistema nervoso, segundo ele o movimento muscular era atribuído a uma substância, os *espíritos animais*⁵, gerada no sangue, agitada por fogo, no coração, e percebida como uma pressão de ar pela glândula pineal (que por sua vez intermediava a relação entre a alma e o corpo)(DESCARTES, 1955).

⁴ Androide, pelo prefixo “andro-”, se refere à forma masculina; enquanto ginoide, toma do grego γυνή, para referir-se à forma feminina, através do prefixo “gin-”.

⁵ A palavra “espírito” aqui tem um caráter material e não religioso nem transcendente. “Espírito”, aqui, é similar ao uso da palavra *spirit* para designar o álcool presente na água durante o processo de produção das bebidas destiladas. Esta explicação tentava justamente mostrar que não é necessária uma alma para as funções corporais, seja em humanos ou demais animais.

Ao se deparar com os autômatos de Vaucanson e com a descoberta do princípio de funcionamento dos músculos, Julien Offray de La Mèttrie (1709-1751) pôde descartar a necessidade de *espíritos animais* para explicar o movimento dos seres vivos. Seu livro de 1748, cujo título pode ser traduzido como “Homem Máquina”, é resultado de suas observações da relação entre doenças físicas e mentais. Observa que existe um grau de complexidade crescente desde o pato ao androide flautista e reflete: a distinção entre humanos e animais não se dá por aquele possuir uma natureza diversa deste, mas a uma mera diferença quantitativa, de algo como roldanas e molas e, sendo assim, o homem seria apenas uma máquina a rastejar perpendicularmente. Este texto traz o humano, também, para a esfera dos autômatos, e o pensamento sendo o resultado da organização complexa da matéria. Assim, se animais são autômatos complexos, humanos são apenas animais mais complexos (WOOD, 2002, 2012):

It [thus] appears that there is but one [type of organization] in the universe, and that man is the most perfect [example]. He is to the ape, and to the most intelligent animals, as the planetary pendulum of Huyghens is to a watch of Julien Leroy. More instruments, more wheels and more springs were necessary to mark the movements of the planets than to mark or strike the hours ; and Vaucanson, who needed more skill for making his flute player than for making his duck, would have needed still more to make a talking man, a mechanism no longer to be regarded as impossible, especially in the hands of another Prometheus. (METTRIE, 1912, p. 140-141)

Em relação ao crescente conhecimento a respeito do funcionamento do corpo humano, a descrição da circulação sanguínea realizada por Willian Harvey (1578 - 1657) em seu texto *Exercitatio Anatomica de Motu Cordis et Sanguinis in Animalibus*, de 1628, teve especial relevância. O movimento do sangue no corpo foi compreendido como um conjunto de tubos, bombas e válvulas, seguindo um comportamento mecanicamente previsível. Neste sentido, Luis XV, para aprimorar a formação de cirurgiões⁶ na França, encomendou de Vaucanson, em torno de 1740, o que seria o seu mais ousado projeto: um modelo didático humano que replicava a circulação sanguínea, respiração, nervos e digestão. Embora tenha trabalhado mais de 25 anos no projeto, a limitação dos materiais disponíveis na época impediu a compleição de seu homem sintético (WOOD, 2002).

⁶ Nesta época os cirurgiões não pertenciam a classe médica, sendo parte da guilda dos barbeiros, e Luiz XV atribuía a morte de seu avô, Luis XIV, à falta de profissionais competentes.

“O Turco”, um suposto autômato jogador de xadrez, representou um enigma exemplar para a hipótese do pensamento mecânico (*Figura 1*). Considerado um jogo de extrema habilidade intelectual, a ideia de uma máquina capaz de derrotar um oponente humano trouxe o debate da AI às cortes e salões do século XVIII. No entanto, para realizar tal feito, o dispositivo dissimulava em seu interior um jogador humano quem, através de imãs, tomava conhecimento da posição das peças no tabuleiro e, por um mecanismo pantográfico, controlava os braços e mãos do boneco. Os segredos do seu funcionamento foram mantidos até após a destruição da máquina por um incêndio no museu que a abrigava, em 1854. E, talvez devido em parte ao artil empregado na concepção deste autômato, a habilidade de jogar xadrez permaneceu até os anos finais do século XX como ápice da genialidade do intelecto humano (MCCORDUCK, 2004, WOOD, 2002).

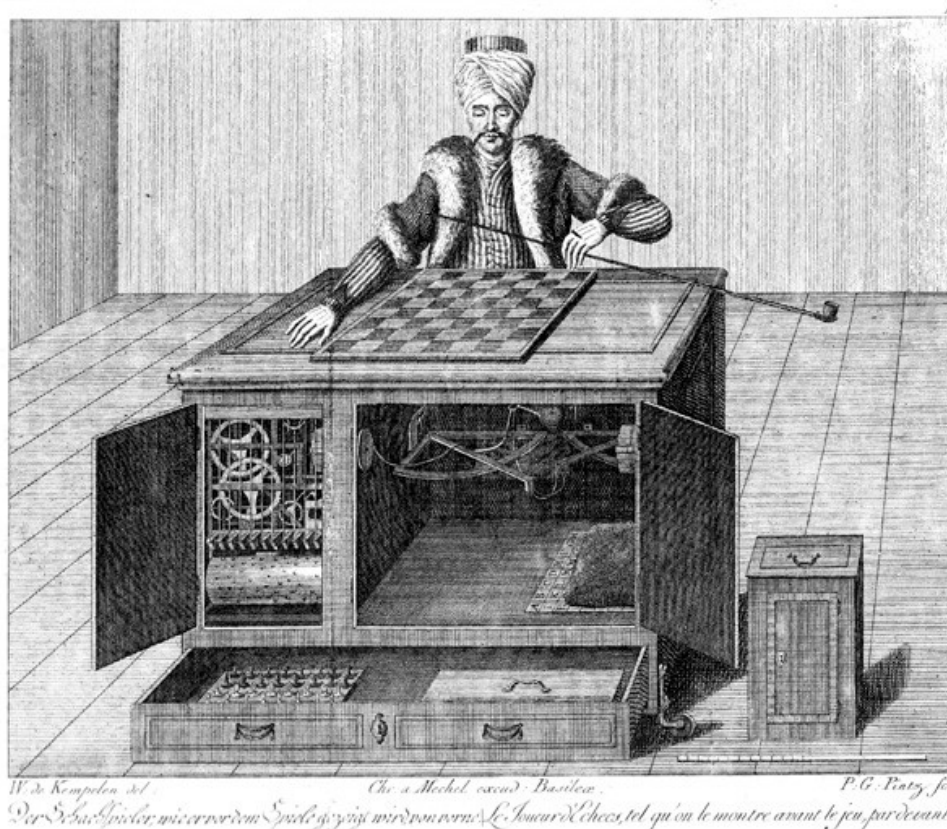


Figura 1: O Turco. Gravura em metal (cobre) de 1783. Fonte: Wikipedia, domínio público. Extraída de WINDISCH, Karl Gottlieb von. **Briefe über den Schachspieler des Hrn. von Kempelen, nebst drei Kupferstichen die diese berühmte Maschine vorstellen.** Basileia: C. von Mechel, 1783. Domínio Público.

2.4 *Pensamento mecânico na Antiguidade e Medievo*

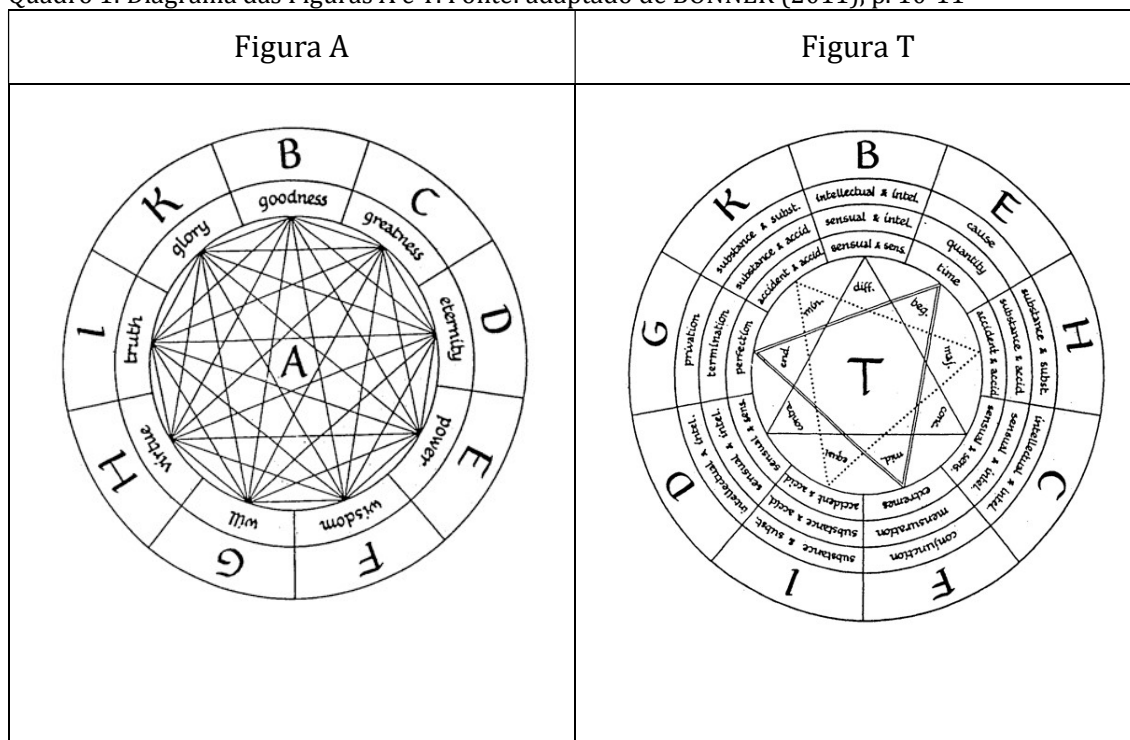
O que a cultura popular mais associa à AI é, obviamente, a imagem de um androide ou robô. Porém o verdadeiro mais antigo ancestral conhecido no campo da IA como, de fato, se realiza hoje é, talvez, o mecanismo de *Antikythera*. Na região da Grécia Antiga, máquinas já estavam sendo produzidas para realizar analogicamente complexos cálculos sobre a posição de planetas, datas de estações do ano e eclipses. Os gregos já utilizavam vapor e reservatórios de água para mover mecanismos, tendo, como exemplo, portas automáticas, uma variedade de clepsidras, relógio com alarme e até mesmo uma máquina de venda automatizada⁷. Mas o mecanismo de *Antikythera*, porém, era operado manualmente. Não encontramos evidência da ideia de vincular a operação de cálculo a uma fonte de energia externa até a modernidade, com Babbage – e isto apenas na forma de projeto. Máquinas de calcular totalmente mecânicas ainda se encontravam em operação há menos de trinta anos no dia a dia de muitas profissões como contador, engenheiro e operador de caixa.

Ramon Llull (1232–1316) foi um controverso filósofo que tentou reproduzir o funcionamento do pensamento através de um mecanismo de anéis concêntricos que realizavam combinações de ideias (Quadro 1). O seu sistema, a *Ars Magna*, é menos interessante do que parece à primeira vista: basicamente tudo o que fazia era permutar diversas letras que representavam ideias escolhidas por Lull por razões religiosas e teológicas. Sua crença era a de que todo conhecimento era composto por combinações de verdades básicas, como átomos de saber. Ele foi fortemente inspirado pela *Sefer Yetzira* ou *Livro da Criação*, a fonte da tradição cabalística. Esta, defendia que todas as coisas existentes eram criadas pelo deus dos judeus através da combinação das 22 letras do alfabeto hebraico. O intuito do seu sistema seria produzir um conhecimento sistemático através do qual a verdade do cristianismo seria evidente e, a seguir, usá-lo para converter os povos islâmicos. A *Ars Magna*, não

⁷ Heron de Alexandria (Séc. I EC) criou uma máquina operada por moedas para ser usada em templos na venda de água cerimonial. Em seu livro *Pneumatica* há a descrição detalhada do mecanismo de diversas máquinas movidas por ar, água e vapor. Ainda antes dele, Ctesibius de Alexandria (Séc. III AEC), já havia usado água e ar para gerar força motriz em relógios e autômatos, mas nenhum texto original foi preservado, restando apenas as descrições de Vitruvius (Séc. I AEC) que, acredita-se, teve acesso aos textos originais. Até meados do século XVII, antes da introdução do pêndulo, mecanismos baseados no relógio de Ctesibius eram os mais precisos para marcar o tempo.

obstante, não encontrou ressonância nem mesmo entre os cristãos, sendo vista com desconfiança pelos acadêmicos e como um atalho para os que não tinham a capacidade de acompanhar as minúcias do saber escolástico⁸ (BONNER, 2007, 2011, CROSSLEY, 2011, SALES, 2011).

Quadro 1: Diagrama das Figuras A e T. Fonte: adaptado de BONNER (2011), p. 10-11



Bonner (2011) nos dá o exemplo da combinação entre **B** e **C** no *Quadro 1*. Na *Figura A* encontramos Bondade e Grandeza, conquanto que na *Figura T* encontramos Diferença e Concordância, e Sensual e Intelectual. Isso permitiria analisar a frase “Bondade tema Grande Diferença e Concordância” aplicado às áreas do Sensual/Sensual, Intelectual/Sensual e Sensual/Intelectual. Combinando essas verdades, acreditava Lull, não se poderia derivar um saber falso. Sendo assim o conhecimento obtido pelo método combinatório seria sempre verdadeiro.

⁸ Escolástica é a filosofia cristã da Idade Média, baseada na preleção (*lectio*) e nas disputas filosóficas (*disputatio*). Aqui se trata da primeira fase da Escolástica (Séc. IX a XII EC), em que se acreditava na harmonia entre fé e razão e coincidência de seus resultados. Neste contexto, a filosofia deve conduzir à fé através da compreensão da verdade revelada, da rejeição das heresias e da defesa contra incredulidade (ABBAGNANO, 2003).

Quadro 2: Terceira Figura. Combinações das Figuras A e T. Adaptado de: BONNER, 2011

BC	CD	DE	EF	FG	GH	HI	IK
BD	CE	DF	EG	FH	GI	HK	
BE	CF	DG	EH	FI	GK		
BF	CG	DH	EI	FK			
BG	CH	DI	EK				
BH	CI	DK					
BI	CK						
BK							

Como Llull não estabelecia distinção entre as combinações de BC e CB nem permitia repetição de letras, como BB ou CC, a *Quadro 2* mostra todas as combinações binárias possíveis das Figuras A e T. Assim, essas combinações formariam uma matriz simétrica, da qual a diagonal principal seria omitida e as repetições ocultadas. Em matemática, denomina-se uma matriz triangular inferior estrita (Quadro 3).

Quadro 3: Obtenção da Terceira Figura

<i>LxC</i>	B	C	D	E	F	G	H	I	K
B	BB	CB	DB	EB	FB	GB	HB	IB	KB
C	BC	CC	DC	EC	FC	GC	HC	IC	KC
D	BD	CD	DD	ED	FD	GD	HD	ID	KD
E	BE	CE	DE	EE	FE	GE	HE	IE	KE
F	BF	CF	DF	EF	FF	GF	HF	IF	KF
G	BG	CG	DG	EG	FG	GG	HG	IG	KG
H	BH	CH	DH	EH	FH	GH	HH	IH	KH
I	BI	CI	DI	EI	FI	GI	HI	II	KI
K	BK	CK	DK	EK	FK	GK	HK	IK	KK

A Quarta Figura, parece mais misteriosa e é composta por três anéis rotatórios concêntricos, nos quais as letras das verdades atômicas estão inscritas. A existência de três círculos possibilitava 252 sequências de tríades de conceitos sem repetições. Em algumas edições o dispositivo era construído em metal ou pergaminho, separado do corpo do livro, para o estudante ir acompanhando os exemplos. Este método, Llull demonstrou em *Rhetorica Nova*, permitiria a criação de sermões inteiros, simplesmente, girando os anéis e produzindo combinações de verdades a serem exploradas no discurso (BONNER, 2007).

2.5 *Pensamento mecânico na Modernidade*

Em 1889, trabalhadores foram contratados para consertar infiltrações no telhado da Universidade de Göttingen, localizada no noroeste da Alemanha. Ao verificarem o ático, encontraram uma estranha máquina com cilindros de bronze e manivelas em madeira. Estava esquecida lá, em um canto, por mais de 100 anos. O mecanismo era o último protótipo da máquina de calcular de Gottfried Wilhelm Leibniz (1646 – 1716). O invento resultou da tentativa de aprimorar o *pascaline*, que era capaz de somar e subtrair, desenvolvido por Pascal alguns anos antes. Sem sucesso, Leibniz acabou por desenhar uma máquina inteiramente nova para ser capaz de processar multiplicações. O projeto iniciou em 1672, completando o primeiro protótipo no ano seguinte, quando o apresentou à *Royal Society of London*. Seu design permitia multiplicar e dividir, mas possuía, porém, falhas que impediam realizar corretamente cálculos com várias operações de *carry* (SCHWARTZ, 2019).

Sendo leitor de Llull, Leibniz, em uma de suas primeiras obras, *Dissertatio de arte combinatoria*, de 1666, expande a ideia de Descartes de que a linguagem possuiria um conjunto básico de elementos primitivos e imagina a possibilidade de um método para combinar esses elementos de modo a produzir saber filosófico, que consiste em ciência, matemática e metafísica. Desde então, em vários textos, Leibniz menciona uma língua artificial, a *characteristica universalis*, que seria constituída por um alfabeto do pensamento – pictográfico e independente das palavras do idioma falado – e diagramas (SCHWARTZ, 2019). Este alfabeto seria ausente de ambiguidade e poderia ser associado mecanicamente, de maneira que sublinha a

influência do método de Lull, através de uma lista de regras lógicas para combinações e recombinações válidas. Ao invés de simples discos concêntricos, as combinações iriam requerer uma máquina mais aos moldes de sua calculadora mecânica.

In this way he was led to imagine that all truths can be deduced from a small number of simple truths by an analysis of the concepts they contain, and that from these in turn, all ideas can be reduced by decomposition to a small number of primitive, undefinable ideas. Thereafter, it would suffice to make a complete enumeration of these simple ideas, the genuine elements of all thought, and to combine them, in order to obtain progressively and by means of an infallible procedure all the complex ideas. In this way, one would construct the *alphabet of human thoughts*, and all derivative notions would be no more than combinations of the fundamental notions, just as the words and sentences of speech are no more than combinations, indefinitely varied, of the 26 letters of the alphabet. (COUTURAT, 1997)

Essa máquina descrita na citação acima por Couturat, denominada por Leibniz *calculus ratiocinator*, é o que Norbert Wiener resgata em seu livro *Cybernetics*, com o qual, pretendia o filósofo, todas as disputas poderiam ser resolvidas simplesmente por meio de cálculos⁹:

Indeed, Leibniz himself, like his predecessor Pascal, was interested in the construction of computing machines in the metal. It is therefore not in the least surprising that the same intellectual impulse which has led to the development of mathematical logic has at the same time led to the ideal or actual mechanization of processes of thought. (WIENER, 1961)

Charles Babbage (1791 – 1871) elevou o conceito de computação mecânica a um nível superior de complexidade com o projeto da Máquina Diferencial, a qual seria capaz de tabular funções polinomiais. O objetivo do projeto seria produzir de forma automatizada tábuas de logaritmos, evitando assim os erros, tanto de cálculo quanto de impressão, comuns nas tabelas então existentes. Antes de poder completar sua construção – abandonando a construção da Máquina Diferencial Nº 1 para se dedicar a um modelo aprimorado, denominado Máquina Diferencial Nº 2, no

⁹ *Quando orientur controversiae, non magis disputatione opus erit inter duos philosophos, quam inter duos Computistas. Sufficiet enim calamos in manus sumere sedereque ad abacos, et sibi mutuo (accito si placet amico) dicere: calculemus.* (LEIBNIZ, 1978, p. 200)

“Quando uma controvérsia aparecer na disputa entre dois filósofos, eles não precisarão argumentar mais que dois contadores [computadores]: basta tomar a caneta e o ábaco e dizer um ao outro, como quem convida a um amigo, vamos calcular!” (Tradução nossa)

Esse otimismo de Leibniz foi ridicularizado por Voltaire em seu *Cândido*, através de seu personagem Dr. Pangloss. Gregory Chaitin argumenta que esta leitura não faz jus à visão de Leibniz, o que este teria em mente é mais próximo a um deus que computa ideias usando uma estratégia de otimização e devolve como resultado o melhor dos mundos possíveis.

meio desse processo –, percebeu que poderia construir uma máquina capaz de realizar qualquer computação desde que alimentada com o conjunto de instruções necessárias (BABBAGE, 1864).

A Máquina Analítica (ver Figura 2), projetada em 1837 e nunca construída, foi o primeiro sistema a utilizar-se de programas armazenados – em papel perfurado similar o usado nos teares mecânicos –, com um setor dedicado à memória, para armazenar as variáveis e resultados, e outro para as operações propriamente ditas. Para a tecnologia de usinagem e fundição da época, o nível de precisão necessário para fazer funcionar uma máquina deste porte estava muito além do que existia. O investimento do governo britânico em suas ideias cessou quando, apesar do gasto do equivalente a dez locomotivas e dez anos de trabalho, Babbage não tinha mais o que mostrar além de um pequeno protótipo que usava para demonstrar as possibilidades do projeto. Além das dificuldades técnicas, sua personalidade não colaborava: ao solicitar o financiamento do governo, atacava, simultaneamente, os dignitários deste governo – a quem cabia decidir sobre o recurso – pela falta de visão ao lhe negar os fundos de que precisava. E, após um desentendimento com o artesão que estava produzindo as partes necessárias para a Máquina Diferencial Nº 1, as peças já produzidas foram derretidas e a construção finalmente interrompida em 1833.

Apesar dos obstáculos, a Máquina Diferencial chegou a ser construída¹⁰, porém não no tempo de Babbage, mas como resultado de um trabalho que começou em 1980 e só concluído em 2002. A construção da Máquina Analítica, no entanto, continua na forma de projeto, tanto pela complexidade do empreendimento como pela notação incomum usada por Babbage nos seus diagramas que torna muito difícil a interpretação dos seus desenhos.

Ada Lovelace (1815 – 1852), filha do famoso poeta, Lord Byron, uma menina que enquanto criança projetava construir um avião, baseado na fisiologia das aves e movido a vapor, e quando adolescente fugiu com o tutor, tornou-se uma matemática

¹⁰ Uma máquina baseada no esquema da Máquina Diferencial de Babbage foi construída por Pehr Scheutz e seu filho Edvard Scheutz entre 1837 e 1853, a qual foi usada para calcular e imprimir tabelas logarítmicas. Estas tabelas eram incompletas, porém. Babbage ficou muito impressionado e recomendou que Pehr Scheutz recebesse uma medalha da Royal Society.

brilhante que anteviu o potencial da Máquina Analítica. Quando ainda jovem, aos 17 anos, foi apresentada a Babbage que, admirado com os interesses matemáticos da jovem, a convidou para uma demonstração de seu protótipo da Máquina Diferencial Nº 1. Apesar de ser apenas uma sétima parte do que seria a Máquina Diferencial, este era então o mais avançado mecanismo de cálculo já criado. Este evento causou uma profunda impressão em Ada que em seus diários descreveu aquele conjunto de manivelas e engrenagens como *thinking machine* e passou, desde então, a acompanhar as ideias de Babbage (KIM, TOOLE, *et al.*, 1999, TOOLE, 1992).

Em 1840 Babbage esteve em Turin, Itália, apresentando o projeto da Máquina Analítica (*Figura 2*), na qual descreveu o mecanismo e apresentou alguns exemplos, o que hoje chamamos de programas de computador. Na audiência um oficial francês, L. F. Menabrea, se mostrou bastante atento e publicou em 1842 um artigo no qual descrevia detalhadamente como estes “programas” seriam implementados no mecanismo. Estes primeiros programas mostravam operações simples como a série dos quadrados, a solução de sistemas de equações, e simples operações algébricas. Lovelace traduziu este texto para o inglês e pretendia publicá-lo. Porém Babbage fez uma sugestão que não a havia ocorrido: ao invés de traduzir, escrever ela mesma um texto sobre este assunto com o qual estava, afinal, tão familiarizada. Após quase dois anos de trabalho e intensa correspondência com Babbage, a tradução foi publicada, com as de Lovelace *Notas* ao final do texto. (TOOLE, 1992)

A Máquina Analítica se distinguia essencialmente das “máquinas de calcular”, uma vez que pretendia não apenas à manipulação de valores, mas de qualquer padrão que pudesse ser representado como símbolos. Neste sentido, em 1840, na correspondência trocada com Babbage, Lovelace revela seu interesse em escrever o jogo *Peg Solitaire* em linguagem matemática. (TOOLE, 1992). Em sua *Nota A*, Lovelace também discute como a música poderia ser descrita cientificamente em relações e que a Máquina Analítica poderia compor elaboradas peças de música em *qualquer grau de complexidade ou extensão* (MENABREA, LOVELACE, 1842).

The bounds of *arithmetic* were however outstepped the moment the idea of applying the cards had occurred; and the Analytical Engine does not occupy common ground with mere “calculating machines.” It holds a position wholly its own; and the considerations it suggests are most interesting in their nature. In enabling mechanism to combine together *general* symbols in successions of unlimited variety and extent, a uniting link is established between the operations of matter and the abstract

mental processes of the *most abstract* branch of mathematical science. (...) Thus not only the mental and the material, but the theoretical and the practical in the mathematical world, are brought into more intimate and effective connection with each other. (MENABREA; LOVELACE, 1842, Nota A.)

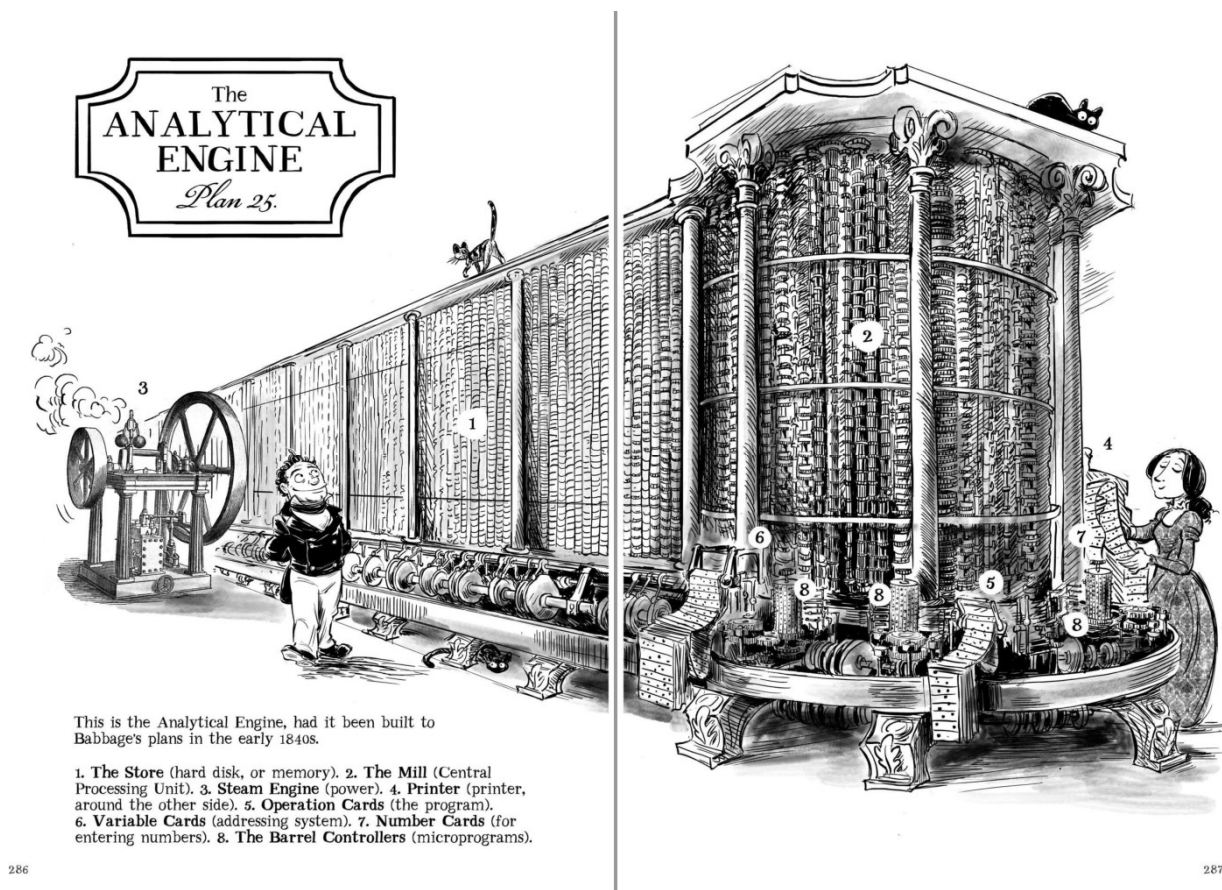


Figura 2: Cartoon em duas páginas representando a Máquina Analítica, a numeração aponta as partes principais, com legenda na própria figura. Fonte: PADUA, Sydney. **The Thrilling Adventures of Lovelace and Babbage**. Randon House: Toronto, 2015. Págs. 386-387.

Ao final de sua *Nota G*, Lovelace fez uma afirmação que, mais de um século depois, Turing acabou contestando, ganhando destacada importância para a área da computação:

The Analytical Engine has no pretensions whatever to originate anything. It can do whatever we know how to order it to perform. It can follow analysis; but it has no power of anticipating any analytical relations or truths. Its province is to assist us in making available what we are already acquainted with. This it is calculated to effect primarily and chiefly of course, through its executive faculties; but it is likely to exert an indirect and reciprocal influence on science itself in another manner. For, in so distributing and combining the truths and the formulæ of analysis, that they may become most easily and rapidly amenable to the mechanical combinations of the engine, the relations and the nature of many subjects in that science are necessarily thrown into new lights, and more profoundly investigated. This is a decidedly indirect, and a somewhat speculative, consequence of such an invention. It is however pretty evident, on general principles, that in devising for mathematical truths a

new form in which to record and throw themselves out for actual use, views are likely to be induced, which should again react on the more theoretical phase of the subject. There are in all extensions of human power, or additions to human knowledge, various collateral influences, besides the main and primary object attained.(MENABREA; LOVELACE, 1842, Nota G.)

2.6 Lógica

A lógica tem origem na Grécia antiga, sendo Aristóteles (384 – 322 AEC) quem realizou a primeira sistematização do tema, seus principais trabalhos que tratam da lógica são o *Analítica Primeira* e *Sobre a Interpretação*. Estas obras foram as mais importantes referências sobre a lógica na antiguidade e permaneceram como a base da disciplina por mais de dois milênios. Os gregos já haviam percebido que promover o convencimento sobre um assunto não é o mesmo que expressar a verdade sobre este, o que atribuiu marcada relevância à lógica, haja vista tratar-se de um povo que dava grande importância ao comércio e à política, sobretudo na democracia ateniense. Em outros termos, os cidadãos da Grécia antiga atentaram para o fato de que nosso raciocínio tende a cometer erros caso não seja devidamente preparado, treinado e temperado contra o mal uso da arte retórica. Desta forma, a lógica se tornou uma ferramenta para a prática da filosofia. A dedução aristotélica, comumente expressa na forma do silogismo, separa o conteúdo do discurso de sua expressão, de acordo com a doutrina peripatética¹¹ de matéria e forma. O que permite analisar a verdade das premissas de maneira dissociada da validade do argumento e, assim, construir uma cadeia dedutiva que preserve a verdade, estruturalmente independente do assunto de que se versa (*Quadro 4*).

¹¹ O termo peripatético refere-se aos seguidores da escola de Aristóteles.

Quadro 4: Exemplo típico de silogismo de Primeira Figura¹², do Modo cujo mnemônico é *Barbara*, as vogais do mnemônico indicam que duas proposições universais afirmativas levam a uma conclusão universal afirmativa.

<p>O silogismo:</p> <p>Todo humano é mortal. (<i>premissa maior</i>)</p> <p>Todo ateniense é humano (<i>premissa menor</i>)</p> <hr/> <p>Logo, todo ateniense é mortal. (<i>conclusão</i>)</p> <p>Pode ser descrito formalmente,</p> <p>Todo B é A.</p> <p>C é B.</p> <hr/> <p>Logo, todo C é A</p> <p>Ou, mais sinteticamente, <i>a</i> indica a cópula:</p> <p>AaB, BaC: AaC</p>
--

Quadro 5: As Quatro Figuras do Silogismo. S: sujeito, P: predicado, M: termo médio. Retomando o exemplo do Quadro 1, temos “humano” como termo médio, “mortal” como predicado e “ateniense” como sujeito.

	Primeira Figura	Segunda Figura	Terceira Figura	Quarta Figura
<i>Premissa Maior</i>	M – P	P – M	M – P	P – M
<i>Premissa menor</i>	S – M	S – M	M – S	M – S
<i>Conclusão</i>	S – P	S – P	S – P	S – P

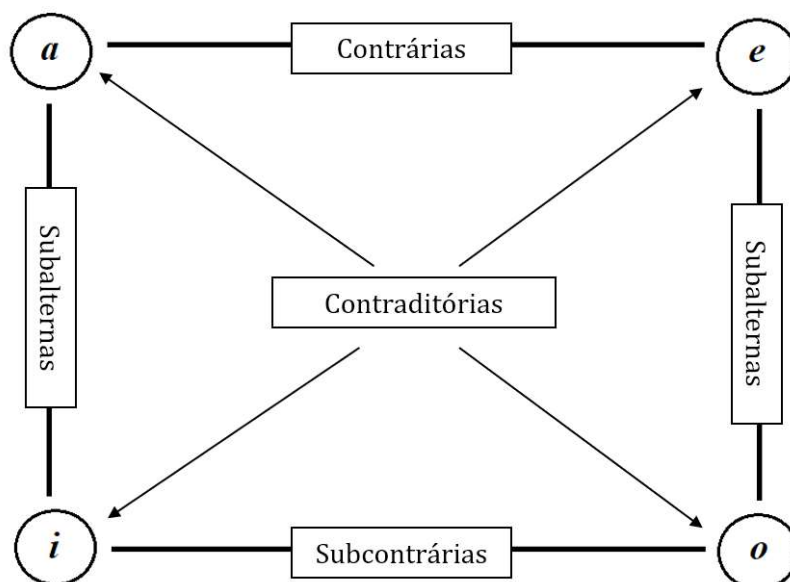


Figura 3: Esquema de relações entre proposições denominado *Quadrado de Oposições* de silogismos.

¹² O termo “figura” refere-se à forma geral do silogismo, em relação ao seu termo médio (Ver *Quadro 5*). O termo médio é o elemento que explica a conexão entre o predicado e o sujeito, no *Quadro 4*, por exemplo, B conecta C a A. O “modo” refere-se ao tipo de proposição segundo as cópulas utilizadas (ver *Figura 3*).

A forma do silogismo admite quatro diferentes cópulas, “Todo é”, abreviado pela vogal *a*, “Nenhum é”, *e*, “Algum é”, *i*, e “algum não é”, *o*. Assim forma-se quadrado de oposições (*Figura 3*).

Usando o Quadrado de Oposições e as Quatro Figuras (*Quadro 5*), temos o total de 4^4 possíveis combinações, das quais apenas 24 são consideradas válidas e foram consolidadas durante a Idade Média nos mnemônicos, conforme *Quadro 6*:

Quadro 6: Mnemônicos medievais para o silogismo¹³.

Figura 1	Figura 2	Figura 3	Figura 4
<i>Barbara</i>	<i>Cesare</i>	<i>Datisi</i>	<i>Calemes</i>
<i>Celarent</i>	<i>Camestres</i>	<i>Disamis</i>	<i>Dimatis</i>
<i>Darii</i>	<i>Festino</i>	<i>Ferison</i>	<i>Fresison</i>
<i>Ferio</i>	<i>Baroco</i>	<i>Bocardo</i>	<i>Calemos</i>
<i>Barbari</i>	<i>Cesaro</i>	<i>Felapton</i>	<i>Fesapo</i>
<i>Celaront</i>	<i>Camestros</i>	<i>Darapti</i>	<i>Bamalip</i>

A lógica aristotélica também discute em quais casos uma figura pode ser convertida na forma de outra. O Filósofo também apresenta provas da validade formal dos silogismos. Desde a antiguidade à modernidade o estudo da lógica incorporou graduais avanços e adquiriu maior sistematização. O trabalho de George Boole (1815 – 1864), *An investigation on the Laws of Thought*, realizou algo comparável ao que Descartes fez à geometria, trazendo a álgebra para os domínios da lógica.

Usando de variáveis *x*, *y* e *z*, que denominou *operandos*, Boole definiu a *multiplicação lógica* pela concatenação destas variáveis, como *xy*, para expressar o que há em comum entre estas duas classes. E o símbolo “+” foi usado para expressar a *soma lógica*, por exemplo, *x+y*, que é a união entre as duas classes *x* e *y*. O número “1” representa “o universo”, e o número “0” representa “nada”, a classe nula (KLEMENT, 2019).

¹³ Conforme foi brevemente explicado acima, *camestres* representa um silogismo válido de segunda figura, desta palavra tomam-se as vogais para produzir a forma *BaA*, *BeC*: *AeC*. Assim o silogismo a seguir é válido, portanto: Todo humano é mortal. Nenhum deus é mortal. Logo, nenhum humano é um deus.

Disto, então, apresenta-se que:

$$xy = yx \quad (1)$$

$$xx = x^2 = x \quad (2)$$

$$xy = 0 \quad (3)$$

$$x + y = 1 \quad (4)$$

$$x + y = y + x \quad (5)$$

$$x(y+z) = xy + xz \quad (6)$$

$$(1 - x) + x = 1 \quad (7)$$

Em (1) notamos a similitude com a comutação da álgebra. Em (2), no entanto, a operação pede pelo que é comum entre duas classes idênticas, que, obviamente, é ela mesma. Em (3), a operação expressa que não há uma área comum entre x e y . Similarmente, em (4) temos que x é o complemento de y . Em (5) aparece a propriedade comutativa da soma, semelhante à da álgebra. Em (6), a propriedade é à distributiva da álgebra. Em (7), Boole nos apresenta, simbolicamente, a lei do terceiro excluído, que pode ser lida como ou x ou *não*- x é verdadeiro (BOOLE, 1854, KLEMENT, 2019).

Retomando o Quadrado de Oposições, temos (*Quadro 7*):

Quadro 7: Representação do Quadrado de Oposições na Álgebra de Boole. O operador v representa uma classe não nula.

Cópula	Expressão	Representação	Representação alternativa
a	Todo X é Y	$xy = x$	$x(1-y) = 0$
e	Nenhum X é Y	$xy = 0$	$x(1-y)=x$
i	Algum X é Y	$xy = v$	$xv = yv$
o	Algum X não é Y	$x(1-y) = v$	$xv = (1-y)v$

Boole desenvolveu, com a lógica simbólica, o que Leibniz chamava de um *calculus ratiocinator*. Nas palavras de Leibniz: “um sistema geral de notação em que todas as verdades da razão pudessem ser reduzidas a um cálculo” (Leibniz *apud* COUTURAT, 1997. Tradução nossa.).

William Stanley Jevons (1835 – 1882) publicou *Pure Logic; or, the Logic of Quality apart from Quantity*, em 1864, no qual aprimora o sistema de Boole. Nesta obra interpretou “+” como união entre as classes:

$$x + x = x \quad (8)$$

Em (8), a expressão pode ser lida como a união de duas classes idênticas. Também dispensou o operando v , ao representar as cópulas i e o como $xy \neq 0$ e $x(1-y) \neq 0$, respectivamente (KLEMENT, 2019).

Além disso, Jevons construiu um ábaco lógico (*Figura 4*), uma máquina com um teclado para entrada, similar ao de um piano (BUCK, HUNKA, 1999), que era capaz de derivar mecanicamente a conclusão a partir das premissas¹⁴.

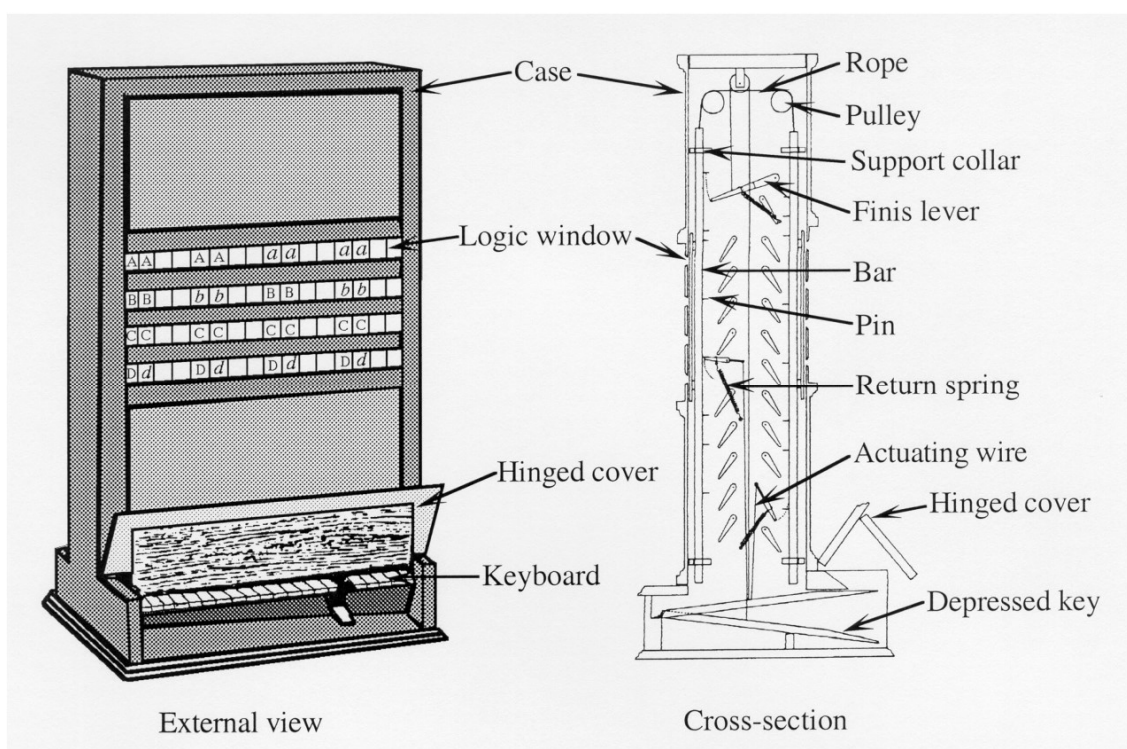


Figura 4: Visão externa e em corte do Ábaco Lógico de Jevons. Fonte: BUCK e HUNKA "W. Stanley Jevons, Allan Marquand, and the Origins of Digital Computing", *IEEE Annals of the History of Computing*, v. 21, n. 4, 1999. p. 22.

Para Gottlob Frege (1848 – 1925), importante matemático e filósofo do século XIX, a lógica tem um caráter mais prescritivo do que descritivo, por não representar a maneira como rotineiramente pensamos. Neste sentido, sua perspectiva é oposta à de Boole, que descreveu a lógica como “Leis do Pensamento”. Frege introduziu

¹⁴ Tecnicamente, um computador digital mecânico. Uma descrição da máquina é dada por Jevons no apêndice da obra *The Substitution of Similars, the True Principle of Reasoning Derived From a Modification of Aristotle's Dictum*, de 1869.

quantificadores para os termos e iniciou a construção de uma sintaxe lógica que denominou *Begriffsschrift*, o que poderia ser traduzido como “notação conceitual” (KLEMENT, 2019). Esta deveria eliminar completamente ambiguidades da linguagem comum e capturar todos os elementos essenciais para as operações lógicas. Com isto pretendia elaborar uma linguagem do trabalho científico, na qual cada termo seria precisamente definido e as regras de inferência seriam derivadas de um pequeno conjunto de axiomas¹⁵. A proposta consistia em eliminar a possibilidade de que erros passassem despercebidos, exigindo que cada verdade demonstrada estivesse fundada apenas em princípios da lógica, sem apelar, em nenhum momento, para a intuição. E, com isto, evitar as ambiguidades da linguagem natural e operações que possuíssem procedimentos e conhecimentos implícitos.

Esta é a continuação do projeto de Leibniz da construção da *characteristica universalis*, indo além da lógica de Boole, que Frege considerava uma implementação imperfeita do *calculus ratiocinator* (MILLÁN, 2020). O *Begriffsschrift* seria uma *língua characterica*¹⁶ e um *calculus ratiocinator* (KLEMENT, 2019). Enquanto a lógica de Boole tratava, separadamente, a lógica categórica da lógica proposicional¹⁷, o sistema de Frege era capaz de representar ambos simultaneamente. Assim Van Heijenoort comenta:

Frege frequently calls Boole’s logic an “abstract logic”, and what he means by that is that in this logic the proposition remains unanalyzed. The proposition is reduced to a mere truth value. With the introduction of predicate letters, variables, and quantifiers, the proposition becomes articulated and can express a meaning. The new notation allows the symbolic rewriting of whole tracts of scientific knowledge, perhaps of all of it, a task that is altogether beyond the reach of the propositional calculus. We now have a lingua, not simply a calculus. (VAN HEIJENOORT, 1967)

O legado da obra de Boole e, principalmente, de Frege, foi tão bem recebido que a lógica aristotélica, que vigorou por mais de dois milênios, foi abandonada no

¹⁵ Esta perspectiva culmina, a partir de 1920, no Positivismo Lógico, tendo como os principais referenciais o Círculo de Vienna e o Círculo de Berlin. Para estes, apenas o conhecimento verificável por observação direta ou por prova lógica possui significado.

¹⁶ Mesmo significado de *characteristica universalis*.

¹⁷ A *lógica categórica* é a que tem o formato da lógica aristotélica, isto é, os símbolos representam categorias: Todo A é B, onde A representa “homem” e B representa “mortal”. Na *lógica proposicional*, um símbolo representa toda uma proposição: H, onde H representa a proposição “Todo homem é mortal”. Uma proposição é sempre ou verdadeira ou falsa.

decorrer de algumas décadas. Não obstante, o programa logicista de Frege mostrou-se incapaz de evitar contradições mesmo na aritmética, como apontado por Bertrand Russell (1872 – 1970) envolvendo a autorreferência no Axioma V, o que ficou conhecido como Paradoxo de Russell-Zermelo¹⁸. Discutir o paradoxo extrapola o escopo deste texto, porém o fato de uma inconsistência ter emergido é importante para o que se segue.

Russell, que perseguia o mesmo objetivo que Frege, propôs, junto a Alfred N. Whitehead (1861 – 1947), a estratificação do sistema lógico. Esta estratégia, então denominada Teoria dos Tipos (*Type Theory*), foi uma forma de evitar que a autorreferência originasse paradoxos. A partir desta nova formulação, um conjunto pode apenas tratar de elementos que estejam em uma camada subalterna. No nível mais inferior se encontra a Lógica de Primeira Ordem (FOL), que apenas pode relacionar elementos. Em um nível acima, está a Lógica de Segunda Ordem (SOL), que pode discutir as relações de conjuntos de elementos. E, assim por diante, lógicas de ordem mais alta vão tratar de conjuntos de conjuntos (*Figura 5*). No livro *Principia Mathematica* (PM), Russell e Whitehead usaram esta estratégia para derivar da lógica a aritmética e, ao final do terceiro volume, deixaram indícios de que toda a matemática poderia ser derivada da mesma forma.

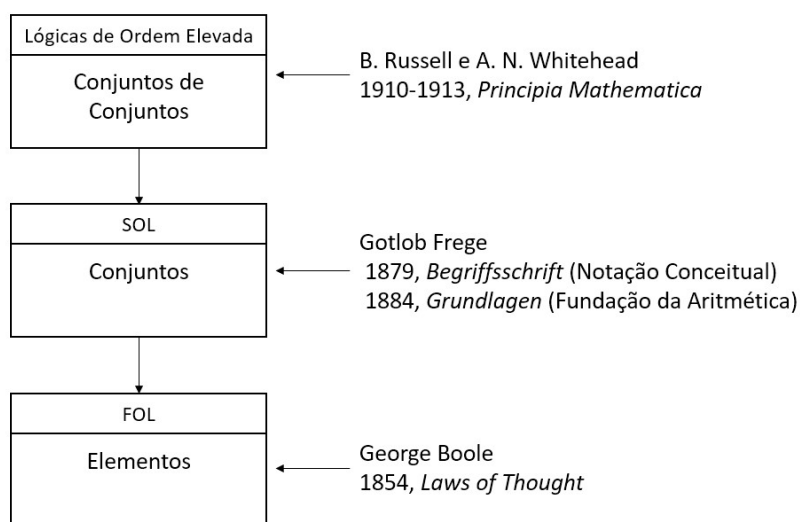


Figura 5: Esquema do desenvolvimento histórico da estratificação hierárquica da Lógica

¹⁸ Esta inconsistência foi descoberta independentemente por Ernst Zermelo (1871 – 1953) – além de outros – alguns anos antes. Para uma introdução básica ao problema ver o artigo HARTNETT, K. "Will Computers Redefine the Roots of Math?", Quanta Magazine, de 19 de maio 2015.

Isso levanta, portanto, duas questões: 1) o PM é consistente? E 2) Derivar toda matemática puramente da lógica é realmente possível? A primeira questão foi, parcialmente, resolvida por Kurt Gödel (1906 – 1978) ao provar o Teorema da Completude, o qual afirma que uma declaração pode ser provada a partir dos axiomas na FOL no PM apenas se é válida (SMULLYAN, 1992). Em outras palavras, se o sistema lógico funciona como uma máquina que pode gerar todas as declarações possíveis através da iteração dos axiomas no qual o sistema é baseado, de tal maneira, esta nunca daria origem a uma declaração falsa. Porém, o processo inverso não está garantido, isto é, dada uma declaração, dizer se esta é válida. Um procedimento ainda precisava ser descoberto, o qual finalmente realizaria o ideal de Leibniz de uma matemática mecanizada (DAVIS, 2001).

Este projeto foi proposto por David Hilbert, em 1928, denominado *Entscheidungsproblem*, ou *problema de decisão*¹⁹, que logo provou-se um objetivo impossível, como foi demonstrado através de Gödel, em 1931, pelos Teoremas da Incompletude²⁰. Até então, acreditava-se que todos os problemas da matemática possuíam solução: ainda que pudessem ser difíceis, ou mesmo que nunca fossem resolvidos, a existência de uma solução não era disputada. Porém, com o resultado de Gödel, soluções negativas também passaram a ser consideradas possíveis.

Paralelamente ao trabalho de Gödel, Alan Turing (1910 – 1954) apresentou um artigo em 1936, intitulado *On Computable Numbers, with an Application to the Entscheidungsproblem*, que fez uso de um recurso que virá a ser muito relevante daí em diante, denominada *máquina de Turing* (PETZOLD, 2008). Seu objetivo era o mesmo que o de Gödel: provar que um sistema era consistente através de um método mecânico e repetitivo. Para isto elaborou um recurso hipotético que utilizaria operações mais simples que pudesse definir, que ficou conhecido como máquina de Turing. Imaginando um matemático com caneta e papel infinito, além de tempo ilimitado para fazer os cálculos, abstraiu essas operações como um *cabecote*, capaz de escrever um símbolo, ler o símbolo, e apagar o símbolo, e o papel

¹⁹ Para uma visão detalhada sobre o Problema de Decisão ver BÖRGER, E., GRÄDEL, E., GUREVICH, Y. **The Classical Decision Problem**. 1st. ed. Berlin, Springer, 1997.

²⁰ Resumindo grosseiramente, nem toda declaração matemática pode ser derivada da lógica. Há declarações que podem ser formuladas no principia, mas dos axiomas no principia não é possível derivar a verdade ou a falsidade.

seria uma *fita* infinita, dividida em células, a qual poderia mover-se para direita ou à esquerda, uma célula por vez. Este dispositivo operaria de acordo com instruções sintáticas, similar ao que hoje chamamos de programas de computador (Tabela 5), que denominou *tabela de instruções* (essa tabela é finita), mas as instruções disponíveis são extremamente limitadas, contendo apenas as operações da fita e do cabeçote. Adicionalmente, há um *registrador de estados*, que armazena o estado da atual máquina na execução da tabela de instruções (o número de estados é finito). A operação da máquina ocorrer em tempo infinito ou terminar após certo tempo (Tabela 1).

Após introduzir este mecanismo hipotético, aplica-se este ao problema de decisão, fazendo uma pergunta mais simples, porém equivalente: É possível se criar uma máquina de Turing que determine, tendo como entrada qualquer Tabela de Instruções, se esta vai parar em algum ponto ou vai continuar executando instruções para sempre. Por exemplo, é fácil ver que nossa Tabela de Instruções para escrever os Números Naturais vai continuar indefinidamente, já que os números naturais são infinitos (Tabela 2).

Em contrapartida, como exemplo, um jogo como *busy beaver* precisa, por definição, terminar em algum momento. Este consiste no desafio de se produzir uma Tabela de Instruções com um tamanho limitado que, ao ser executada, imprima a maior quantidade possível de “1” na Fita. Porém não pode ser um ciclo que imprima a mesma coisa indefinidamente, é necessário que, em algum momento, pare e nos apresente o resultado obtido, isto é, passe ao estado de *halt*.

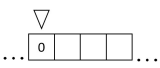
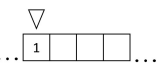
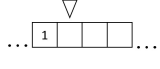
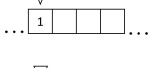
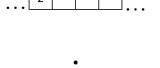
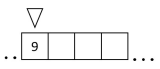
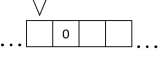
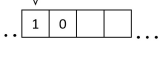
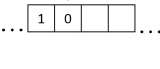

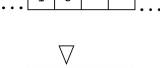

A execução, cabe notar, pode demorar uma quantidade de tempo proibitiva, mesmo para problemas aparentemente simples como o jogo busy beaver²¹. Seria extremamente conveniente se fosse possível tomar todas as máquinas de Turing com até n linhas de instruções e descartar aquelas que entram em ciclos intermináveis. No entanto, a conclusão do artigo de Turing foi que tentar isto para um caso geral levaria a uma contradição.

²¹ De fato, a complexidade do jogo busy beaver cresce mais rapidamente que a função de Ackerman.

Tabela 1: Tabela de Instruções de uma máquina de Turing. Este exemplo imprime todos os números naturais em ordem crescente. Adaptado de PETZOLD, C. **The Annotated Turing: A Guided Tour Through Alan Turing's Historic Paper on Computability and the Turing Machine**. Indianapolis, John Wiley & Sons, 2008.

Estado inicial	Símbolo lido	Operações	Estado final
α	Nenhum	Imprimir 0	β
β	1	Imprimir 1	γ
	2	Imprimir 2	γ
	3	Imprimir 3	γ
	4	Imprimir 4	γ
	5	Imprimir 5	γ
	6	Imprimir 6	γ
	7	Imprimir 7	γ
	8	Imprimir 8	γ
	9	Imprimir 9. Mover para esquerda.	β
	Nenhum	Imprimir 1	γ
γ	Nenhum	Mover para esquerda.	β
	Qualquer	Mover para direita.	γ

Tabela 2: Exemplo de execução de uma máquina de Turing. Notar que a máquina continua indefinidamente.

Passo	Operação	Fita	Estado final
1	Imprimir 0		β
2	Imprimir 1		γ
3	Mover para Direita		γ
4	Mover para Esquerda		β
5	Imprimir 2		γ
⋮	⋮	⋮	⋮
25	Mover para Esquerda		β
26	Imprimir 0 e Mover para Esquerda		β
27	Imprimir 1		γ
28	Mover para Direita		γ
29	Mover para Direita		γ
30	Mover para Esquerda		β
31	Imprimir 1		...

Uma das consequências mais evidentes da formalização do conceito de algoritmo realizada por Turing foi a efetiva criação de máquinas computadoras, os recursos e justificativas que faltaram a Babbage, no século XX abundariam em decorrência da Segunda Guerra Mundial. A vantagem estratégica decorrente da quebra da encriptação das mensagens do Eixo e, posteriormente, a pesquisa da Bomba Atômica e Bomba H, tornou uma necessidade computadores mais rápidos e com mais armazenamento.

Em 1943, Warren McCulloch (1898 – 1969) e Walter Pitts²² (1923 – 1969) publicaram um artigo que ecoava influências do PM, do *On Computable Numbers*, mas também do trabalho de Ramón y Cajal (1852 – 1934) sobre Sinapses neuronais. O título *A logical calculus of the ideas immanent in nervous activity* é bem audacioso ao considerar o cálculo lógico como uma propriedade imanente ao sistema nervoso, que equivale dizer que o cérebro seria um computador digital. A influência de Leibniz perpassa todo seu trabalho, como que encontrando no cérebro humano a corporificação do *calculus ratiocinator*. De forma simplificada, os pesquisadores entenderam que o neurônio é a unidade mínima do sistema nervoso e, a partir desta unidade, elaboraram uma teoria de como o cérebro realiza o processamento de informações, que no entendimento deles significaria implementar algo equivalente a máquinas de Turing.

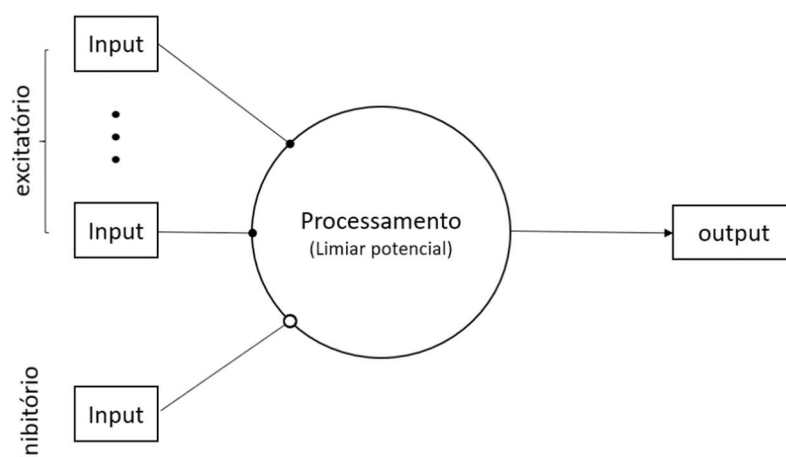


Figura 6: Esquema funcional de neurônio no artigo de McCulloch-Pitts. Linhas terminadas em pontos indicam um input excitatório e a linha terminadas em um laço indica um input inibitório.

Ramon y Cajal, em 1880, formulou a hipótese de que neurônios eram unidades discretas, as quais recebiam sinais pelos dendritos e transmitiam pelos axônios (O'BRIEN, 2006). E também notou que havia uma pequena distância entre as células, dendritos e axônios não estavam diretamente conectados²³. No modelo de

²² Pitts tem uma história de vida muito interessante, abandonou a escola ainda criança para trabalhar, e nas horas vagas leu os três volumes do PM. Escreveu a Russell apontando erros na obra quando tinha apenas 12 anos de idade. Aos 15 anos de idade fugiu de casa para assistir as aulas de Russell que então estava lecionando em Chicago. Foi ali que conheceu McCulloch, eventualmente indo morar com a família dele.

²³ esta distância foi, em 1897, denominada “sinapse”, pelo neurofisiologista Charles Sherrington, que do grego συνάψις significa “conjunção”.

Mcculloch-Pitts, um neurônio seria uma célula que pode ter um único *output* de “0” ou “1”, o qual seria o resultado de processar diversos canais de *input* (Figura 6).

Um input pode ser excitatório ou inibitório. Se um input inibitório estiver ativo o neurônio não irá disparar, isto é, o sinal será “0”. Um neurônio pode ter apenas um input inibitório. Em contrapartida, é necessário ultrapassar um certo limiar potencial de inputs excitatórios para que o neurônio produza o “1”. Este limiar potencial é chamado *threshold value* do neurônio. As operações lógicas como conjunção, disjunção, negação podem assim ser descritas pela combinação destes neurônios.

Com este sistema qualquer função lógica do cálculo proposicional poderia ser representada recursivamente por alguma rede neural a partir das quatro formas básicas (Figura 7). Assim, há uma máquina de Turing capaz de computar qualquer função realizada em uma rede desses neurônios (embora a conversa não seja verdadeira).

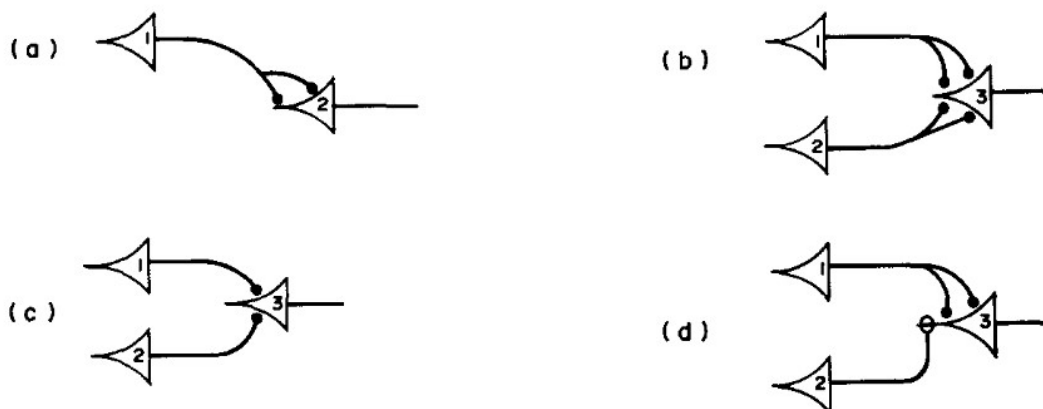


Figura 7: a) Identidade, b) *ou* lógico, c) *e* lógico, d) negação. Linhas terminadas em pontos indicam um input excitatório e a linha terminadas em um laço indica um input inibitório. Fonte: MCCULLOCH, W. S., PITTS, W. "A logical calculus of the ideas immanent in nervous activity", **The Bulletin of Mathematical Biophysics**, v. 5, n. 4, p. 115–133, 1943.

Uma argumentação interessante é oferecida pelo filósofo Gualtiero Piccinini(2004), a qual resumida afirma que, ao mostrar a redução de uma rede neural a uma máquina de Turing, conecta-se as capacidades do cérebro às limitações da cognição humana. Isto é, se de fato as operações da cognição humana são reduzidas às operações que os neurônios implementam, e como estes implementam operações que não ultrapassam o poder de uma máquina de Turing, segue-se que o limite da matemática reside na limitação do cérebro humano.

What we thought we were doing (and I thought we succeeded fairly well) was treating the brain as a Turing machine... The important thing was, for us, that we had to take a logic and subscript it for the time of occurrence of a signal (which is, if you will, no more than a proposition on the move). This was needed in order to construct theory enough to be able to state how a nervous system could do anything. The delightful thing is that the very simplest set of appropriate assumptions is sufficient to show that a nervous system can compute any computable number. It is that kind of a device, if you like a Turing machine. (McCulloch *apud* PICCININI, 2004)

Esta descrição do cérebro não ganhou muitos adeptos na neurofisiologia, mas atraiu a atenção de John von Neumann (1903 – 1957), que ficaria conhecido por desenvolver o modelo de computadores digitais que até hoje nos baseamos. Sua interpretação de circuitos eletrônicos, como equivalentes do sistema nervoso de organismos biológicos, via nos relés e válvulas algo que poderia funcionar como neurônios. E, de fato, válvulas normalmente eram utilizadas para amplificar um sinal nos circuitos analógicos, mas poderiam ser usados de tal maneira que operassem logicamente. Foi o trabalho de Claude Shannon (1916 – 2001), publicado em 1948, intitulado *A Mathematical Theory of Communication*, que permitiu o uso de circuitos digitais para realizar computações sem as degradações associadas aos circuitos analógicos. Conectando válvulas entre si como os neurônios de McCulloch-Pitts, von Neumann poderia criar um computador em que não apenas os números nos dados são entidades abstratas, mas também a operação – não dependendo mais de conectar cabos fisicamente para se realizar um novo procedimento – poderia ser representada por números (NEUMANN, 1993).

2.7 Pioneiros

No ano de 1956 um interessante *workshop* aconteceu em Dartmouth College, o qual reuniu grande parte das pessoas consideradas os pais fundadores da Inteligência Artificial: John McCarthy, Marvin Minsky, Nathaniel Rochester e Claude Shannon, Ray Solomonoff, Allen Newell e Herbert A. Simon, dentre outros. O que unia esses nomes, comenta Pamela McCorduck (MCCORDUCK, 2004), era: *a premissa de que cada aspecto do aprendizado e cada característica da inteligência podem, a princípio, ser precisamente descritos que uma máquina pode ser construída para simulá-los* (MCCARTHY, MINSKY, *et al.*, 2006. Tradução nossa.).

Dentre os temas abordados neste evento apareceram a linguagem natural, a abstração, a criatividade, as redes neurais e até a possibilidade de se escrever um programa que simulasse a mente humana.

An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. (MCCARTHY, MINSKY, *et al.*, 2006)

McCarthy acreditava que unir todas as pessoas que estavam interessadas no assunto em um ambiente propício no qual pudessem se dedicar integralmente levaria a um rápido avanço. No entanto, a ideia de um workshop não funcionou como esperado. Cada cientista estava por demasiado apegado a suas próprias ideias e os encontros eram menos regulares do que o planejado:

Anybody who was there was pretty stubborn about pursuing the ideas that he had before he came, nor was there, as far as I could see, any real exchange of ideas. People came for different periods of time. The idea was that everyone would agree to come for six weeks, and the people came for periods ranging from two days to the whole six weeks, so not everybody was there at once. It was a great disappointment to me because it really meant that we couldn't have regular meetings. (McCarthy *apud* MCCORDUCK, 2004)

2.8 Conclusão e Discussão

A história que rapidamente atravessamos, desde os mitos sobre seres construídos por deuses ou sábios que talvez adentrassem o submundo da bruxaria, até a decisão de se investir, de fato, no projeto da modernidade de se trazer à luz a razão mecanizada, mostra, ao nosso ver, uma busca do humano por sua própria essência, a qual não é mais que um reflexo aberrante de sua própria imaginação.

O deus Hefesto está morto, com toda a cultura que o criou, mas sua arte nunca esteve mais viva. O espírito do deus dos vulcões, que produzia, na era do bronze, as armas e demais equipamentos dos deuses e heróis, anima as forjas atuais que produzem *waffers* de silício para os equipamentos que sustentam todo nosso mundo, de robôs dançantes a drones militares. Os gregos acreditavam que talvez as estátuas estivessem vivas. Cada vez mais vivemos em um mundo com objetos que têm uma estranha aparência de estarem vivos: nos observam continua e atentamente, respondem a nossos gestos, nossa voz, nos consolam e distraem em uma época de isolamento e, para muitos, solidão. Cada vez mais a travessia do *uncanny valley* se torna mais atraente.

A modernidade viu a emergência de uma visão mecânica frente à concepção mística do humano. O que é mecânico pode ser compreendido, imitado e reproduzido em série. Membros do corpo, que se movem através dos nervos, eram similares às figuras que adornavam relógios; os órgãos que fazem o sangue fluir, que processam a digestão e a respiração estavam sendo incorporados em autômatos. Se tudo isso estava sendo imitado, então era um passo natural que as operações da mente passariam pelo mesmo escrutínio. Sendo primeiro engendradas por meio de engrenagens e molas, posteriormente, de relés, válvulas e transistores.

Talvez a lógica não tenha nos dado o poder sobre o conhecimento da mesma forma que acreditávamos. Descobrimos, neste processo, que não somos seres tão racionais quanto Aristóteles supunha. E que as supostas distinções entre razão e não razão – emoção, paixão, instinto, pulsão – que sustentaram sistemas de uma legião de filósofos ao longo da história, corresponde cada vez menos ao que encontramos e conhecemos a respeito de nossos processos cognitivos. A realidade é infinitamente mais cheia de nuances, de tal forma que a análise combinatória de Lull – ou talvez o mecanismo de Lagado²⁴ – pode representar mais a dinâmica interna de nossa cognição do que são capazes as regras da lógica formal: a maquinaria da mente humana parece ser mais bem capturada pela algazarra de redes neurais que pela organização sistemática e hierárquica dos modelos lógicos. Porém, ainda hoje, uma hipótese como esta permanece muito aquém de ser explorada: sem uma teoria de aprendizagem e inteligência que unifique a pesquisa na AI, ela se fragmenta em inúmeras perspectivas e modelos parciais incompatíveis entre si.

²⁴ A máquina de pensar através de combinações simbólicas da Academia de Lagado que foi parodiada por Swift em *As viagens de Gúliver*.

3 APRENDIZAGEM EM MÁQUINAS: ORGANISMOS BIOLÓGICOS E INTELIGÊNCIA ARTIFICIAL

However, since computer time and memory still is a limiting factor, the non-genetic patterns of each numeric symbioorganism are constructed only when they are needed and are removed from the memory as soon as they have performed their task. This situation is in some respects comparable to the one which would arise among living beings if the genetic material got into the habit of creating a body or a somatic structure only when a situation arises which requires the performance of a specific task (for instance a fight with another organism), and assuming that the body would be disintegrated as soon as its objective had been fulfilled.

Barricelli, 1961

3.1 Introdução

Leslie Valiant (2013), em seu livro *Probably Approximately Correct*, oferece uma interpretação que visa compreender a evolução biológica como um processo computacional e, através dessa perspectiva, produzir uma descrição precisa de seu funcionamento interno. Nosso conhecimento atual sobre a evolução, como acontece com diversos campos do conhecimento, está cheio de lacunas. Novos dados poderiam preencher algumas dessas lacunas, e também podem apontar fraturas em nossas atuais teorias.

Embora a Teoria de Darwin forneça um esboço geral, e seja considerada essencialmente correta, ela carece de uma análise quantitativa, uma descrição mecanicista que possibilitaria uma simulação computacional bem-sucedida do processo evolutivo. Tentativas de simular coisas vivas são descritas desde que os computadores se tornaram disponíveis por volta de 1950, entretanto, com poucas exceções quase todas falharam em replicar características semelhantes às encontradas na natureza.

Valiant argumenta que o processo oculto na evolução é algum tipo de algoritmo de aprendizagem, que engloba, ao mesmo tempo, generalidade e simplicidade suficientes. Generalidade, pelo fato de conferir adaptabilidade a um ambiente um tanto incerto, e simplicidade, porque é obrigado a dar o produto de seu cálculo em tempo real. Ele nomeia essa classe de algoritmos de “ecoritmos”, estes não são codificados por um programador de uma teoria, mas emergem da interação de um mecanismo de aprendizagem e seu ambiente. Se esta ideia é correta, a partir

da conjectura Church-Turing (COPELAND, 2019), tais interações poderiam ser descritas como computação e, neste sentido, a Ciência da Computação seria o campo que, por excelência, deveria se dedicar a desvendar os segredos da natureza.

Pretendemos neste trabalho explorar a relação entre a aprendizagem na Inteligência Artificial (AI), e a aprendizagem em organismos biológicos, porém, em um sentido de mais baixo nível ou fundamental. Inicialmente vislumbramos apontar para uma delimitação do termo *aprendizagem* capaz de capturar apenas o que acreditamos possa ser a essência do conceito. Partimos do questionamento sobre como um organismo genérico pode aprender sobre seu ambiente para aumentar o seu sucesso. Neste sentido, não estamos preocupados com mecanismos de alto nível, como memória ou atenção, nem com o funcionamento do tecido neural de seus cérebros, mas com o processo mais simples que permite a um organismo extrair informações específicas do meio, potencializando ações que o beneficiem.

A preocupação com uma definição mais instrumental do conceito de aprendizagem nasceu de sentirmos uma necessidade, de ordem filosófica, de esclarecer certas terminologias, sobretudo quando o uso parece permeado de conotações impregnadas de idealizações que, ao nosso ver, obscurecem o entendimento dos fenômenos.

Para perseguir nosso objetivo, traçamos uma rota a qual iniciamos nos voltando para os mecanismos de evolução. Este trajeto se justifica por duas razões, em primeiro lugar porque entendemos que, qualquer que seja a noção de aprendizagem, é fruto da seleção natural. É um pressuposto através deste trabalho que, qualquer capacidade que considerarmos altamente desenvolvida na espécie humana, tem, atrás de si, uma série de versões e adaptações que remontam ao primeiro organismo vivo. Em segundo lugar porque o próprio mecanismo de evolução nos parece ser, ele mesmo, uma forma de aprendizagem. Este aspecto nos instiga, pois sua elucidação nos ajuda a responder outras questões de ordem científica, propiciando melhor compreensão da história evolutiva. Também favorece o esclarecimento de problemas filosóficos, pois avança na questão sobre como a mente chegou a existir e qual a finalidade da consciência. Por fim, o escrutínio dos mecanismos evolutivos traz contribuições tecnológicas, propiciando a elaboração de melhores sistemas para o desenvolvimento de produtos e serviços na indústria.

Nossa abordagem metodológica consistiu em traçar uma análise reflexiva a respeito da evolução e aprendizado, sob três eixos 1) a possibilidade de se conhecer o mundo, se aprender sobre ele e inferir seu funcionamento 2) a ideia de evolução natural e seu impacto na visão científica de mundo e 3) as consequências epistemológicas da simulação de aspectos da evolução.

Nossa análise usa como literatura de referência o livro de Leslie Valiant, *Probably Approximately Correct: Nature's Algorithms for Learning and Prospering in a Complex World*, de 2013 e a obra *Proving Darwin: making Biology Mathematical*, de Gregory Chaitin, de 2012, em Matemática e Ciência da Computação. Além de *The Growth of Biological Thought: diversity, evolution and Inheritance*, de Ernst Mayr, de 1982, no âmbito da Biologia. Nos amparamos também nos artigos de Nils Aall Barricelli, sobretudo *Numerical testing of evolution theories, Part I*, de 1962. O livro de divulgação científica de Richard Dawkins, *The Blind Watchmaker*, de 1985, sobre evolução, e o de George Dyson, sobre a história do computador digital, *Turing's Cathedral: the origins of the digital universe*, de 2012, servem de base para contextualizar alguns questionamentos, sobretudo os primeiros passos na simulação de organismos vivos.

3.2 A Inteligibilidade do Real

O pensador grego Heráclito de Éfeso (aprox. 535 – 475 AEC) acreditava que a *ordem do mundo* estava de acordo com o *Logos*. Logos tem, nos textos do filósofo, um significado que parece se localizar na intersecção entre razão, palavra, discurso e proporção, expressando um princípio formador do mundo que orienta tudo o que existe (ABBAGNANO, 2003, JOHNSTONE, 2016).

This logos holds always, but humans always prove uncomprehending, both before hearing it and when they have first heard it. For though all things come to be [or, happen] in accordance with this logos, humans are like the inexperienced when they experience such words and deeds as I set out, distinguishing each in accordance with its nature and saying how it is. But other people fail to notice what they do when awake, just as they forget what they do while asleep. (Heraclitus, Fr. 1 *apud* JOHNSTONE, 2016.)

O filósofo para quem tudo flui, parece ver na linguagem um meio de se parar o ser no tempo, para permitir perscrutá-lo. A linguagem, tal qual uma rede lançada sobre todas as coisas, é o que permite tentar apreender o real, que está sempre nos

escapando. Segundo uma perspectiva heraclitiana, apesar da constante inacessibilidade ao real, em sua totalidade e completude, nos aproximamos dele por meio do discurso: com a palavra fazemos a ponte entre o que podemos saber, que está sempre se transformando, e o mundo, que, também, está em contínua mudança

Em um sentido diferente, Platão (aprox. 428 – 348 AEC) – seguindo Pitágoras (570 – 495 AEC) – dava um grande valor à geometria e ciências matemáticas como ferramentas para desenvolver a capacidade de compreender a verdade. A tradição conta que na entrada da Academia havia a inscrição que dizia algo como “não entre quem não for geômetra”. Não há nenhuma evidência de que isso seja verdade, mas retrata a importância do saber geométrico e matemático na ontologia da antiguidade. O mundo, tal como na analogia da caverna, é apenas sombra, mas existe um real do qual as sombras são projeções, e este é acessível à razão.

A mesma crença, de que o mundo é inteligível e manifesta ordem, ainda ecoa, quase dois milênios depois de Platão, nas palavras de Galileo Galilei (1564 – 1642), em 1623:

Philosophy is written in this grand book, the universe, which stands continually open to our gaze. But the book cannot be understood unless one first learns to comprehend the language and read the letters in which it is composed. It is written in the language of mathematics, and its characters are triangles, circles, and other geometric figures without which it is humanly impossible to understand a single word of it; without these, one wanders about in a dark labyrinth. (GALILEI, 1957. p. 238.)

Descartes (1596 – 1650) imagina um gênio enganador que falsifica todas as impressões, em cada um de seus sentidos, que alguém poderia ter da realidade. O que nos salvaria do engano é Deus não ser malevolente (DESCARTES, 1955). Esta perspectiva coloca, dessa maneira, o saber de bases empíricas sobre uma fundação muito frágil. Embora as coisas do mundo, que denomina *res extensa*, pudessem ser mensuradas, analisadas, divididas indefinidamente, combinadas em novas formas, a sua realidade e a realidade das leis que as governavam estavam construídas sobre a base da fé.

Immanuel Kant (1724 – 1804) chama de “escândalo para filosofia e para a razão”, o fato de a existência das coisas no mundo até então não ter sido provada.

It still remains a scandal to philosophy and to human reason in general that the existence of things outside us (...) must be accepted merely on faith, and that if anyone thinks good to doubt their existence, we are

unable to counter his doubts by any satisfactory proof. (KANT, 1998, p.121.)

A nossa compreensão do mundo estaria limitada pelas nossas formas de apreender o real, que para Kant, tendo sido profundamente influenciado por Newton, seriam o tempo e o espaço. No entanto, dando um passo além de Kant, podemos pensar nos diversos saberes inatos dos quais somos dotados como algo equivalente. *A coisa em si*, desde então, está além da capacidade do entendimento humano.

Elaborando ainda mais a questão da inacessibilidade do real, o discípulo de Franz Brentano (1838 – 1917), Edmund Husserl (1859 – 1938) coloca, assim, o mundo *entre parêntesis*, como uma aparição que se revela à consciência. Falar da existência do mundo é, para ele, ingênuo (DA SILVA, 2005).

I am now no longer a human Ego in the universal, existentially posited world, but exclusively a subject for which this world has being, and purely, indeed, as that which appears to me, is present, to me, and of which I am conscious in some way or other, so that the real being of the world thereby remains unconsidered, unquestioned, and its validity left out of account (HUSSERL, 2013).

Em sua filosofia, nem sujeito nem objeto existem independentemente: a única realidade que temos acesso é a relação entre o sujeito que percebe e o objeto percebido, mas nunca a realidade do objeto em si mesmo. E esta relação é mediada pela consciência do sujeito, pela maneira que esta se volta para o objeto, descortinando apenas parte de suas possibilidades de manifestação. Desta maneira, a Filosofia da Ciência de Husserl considera que o saber do mundo natural está limitado a hipóteses – diferentemente da lógica e matemática –, verdades universais e absolutas nunca seriam alcançáveis neste domínio.

1.1.1. UM CAMINHO PARA SE ENTENDER APRENDIZAGEM

Como um dos elementos centrais de nosso estudo são os algoritmos de aprendizagem, inicialmente apresentaremos a noção adotada aqui a esse respeito. A aprendizagem representa uma grande parte do conjunto de ideias que estão por detrás do conceito de inteligência. Uma questão que merece exame minucioso é se há, de fato, uma possível intersecção entre inteligência ou aprendizado em organismos biológicos e AI. Por razões pragmáticas, guardamos este escrutínio para

investigações futuras e trabalhamos com a hipótese de que há possível intersecção entre aprendizado ou inteligência biológica e artificial. E, se queremos explorar o que há de comum entre organismos biológicos e a AI, se faz necessário explorar um pouco o que temos a intenção de dizer através do uso do termo *aprendizagem*. E, neste contexto, trata-se de uma sequência de passos que tornam um agente, indivíduo ou conjunto de indivíduos, através de exclusivamente interagir com o meio, ou mediado por alguma forma de tutoria, mais eficiente em sua interação. Apesar de o uso comum da palavra aprendizagem ser muito mais rico e cheio de sentidos, aqui sequer exigiremos que o sujeito da aprendizagem desenvolva entendimento sobre o objeto aprendido, apenas um mensurável acréscimo na qualidade da performance será suficiente para caracterizarmos como uma aprendizagem.

Outrossim, estamos preocupados com o tipo de melhoria de performance que emerge espontaneamente em um sistema. Por exemplo, consideramos aprendizagem o que as formigas fazem ao mapear seu território encontrando o melhor trajeto entre elas e as fontes de alimento, igualmente consideramos aprendizagem quando uma rede neural, após ser submetida a exemplos, extrai padrões que a permitem classificar, com certo grau de sucesso, amostras similares.

3.3 Evolução *Biológica*

Existem ideias que possuem um grande impacto em como vemos o mundo, ideias a partir das quais tudo precisa ser realinhado para fazer sentido. Um grande exemplo disso parece ser a ideia de evolução através de seleção natural. Está é, talvez, a melhor ideia que alguém já teve, comenta o filósofo da mente Daniel Dennett:

If I were to give an award for the single best idea anyone ever had, I'd give it to Darwin, ahead of Newton and Einstein and everyone else. In a single stroke, the idea of evolution by natural selection unifies the realm of life, meaning, and purpose with the realm of space and time, cause and effect, mechanism and physical law. But it is not just a wonderful scientific idea. It is a dangerous idea (DENNETT, 1995. p. 21.)

Essa ideia é descrita por Charles Darwin (1809 – 1882) como o resultado da luta pela vida, na qual mesmo uma pequena variação, ao longo do tempo, pode

conferir vantagem significativa aos indivíduos que a possuem, levando estes a sobreviverem e deixarem descendentes, perpetuando a variação.

All these results, as we shall more fully see in the next chapter, follow from the struggle for life. Owing to this struggle for life, any variation, however slight, and from whatever cause proceeding, if it be in any degree profitable to an individual of any species, in its infinitely complex relations to other organic beings and to external nature, will tend to the preservation of that individual, and will generally be inherited by its offspring. The offspring, also, will thus have a better chance of surviving, for, of the many individuals of any species which are periodically born, but a small number can survive. I have called this principle, by which each slight variation, if useful, is preserved, by the term of Natural Selection, in order to mark its relation to man's power of selection. We have seen that man by selection can certainly produce great results, and can adapt organic beings to his own uses, through the accumulation of slight but useful variations, given to him by the hand of Nature, But Natural Selection, as we shall hereafter see, is a power incessantly ready for action, and is as immeasurably superior to man's feeble efforts, as the works of Nature are to those of Art. (DARWIN, 2009. p.50.)

Apesar de enunciado de forma simples, a formulação precisa da teoria da seleção natural permanece com várias lacunas. Ernst Mayr (1982), em seu livro *The Growth of Biological Thought*, apresenta uma organização lógica da teoria da seleção natural. Seu modelo parte de cinco fatos, dos quais faz três inferências, que resumimos aqui (*Quadro 8*):

Quadro 8: Lógica da Seleção Natural segundo Mayr (1982)

<p>Fato 1: <i>Todas as espécies têm o potencial de incrementar a população exponencialmente</i></p> <p>Fato 2: <i>As populações são estáveis no tempo</i></p> <p>Fato 3: <i>Os recursos naturais são limitados e estáveis</i></p> <p>Inferência 1: <i>deve, portanto, haver uma competição entre os indivíduos, resultando que, em cada geração, apenas uma pequena parte sobrevive</i></p> <p>Fato 4: <i>Os indivíduos em uma população apresentam grande variabilidade.</i></p> <p>Fato 5: <i>Muito dessa variação é hereditária.</i></p> <p>Inferência 2: <i>O processo de seleção natural é fruto da desigualdade na sobrevivência.</i></p> <p>Inferência 3: <i>a uma mudança gradual da população, eventualmente produz novas espécies</i></p>

Fato 1: Todas as espécies têm o potencial de incrementar a população exponencialmente, que é o princípio de Malthus. Isto significa que, não havendo nenhuma restrição como predação ou escassez de recursos, uma espécie continuaria a aumentar a quantidade de indivíduos indefinidamente, o que não se observa na realidade. Isto nos leva ao seguinte fato:

Fato 2: As populações são estáveis no tempo. Descontando as flutuações sazonais e pequenas variações, as populações não manifestam o comportamento de continuar aumentando indefinidamente, o que nos conduz ao terceiro fato:

Fato 3: Os recursos naturais são limitados e estáveis. O que nos leva, portanto, à primeira inferência:

Inferência 1: Uma vez que mais indivíduos são produzidos do que o ambiente é capaz de suportar com seus recursos e, não obstante, a população continua estável, deve, portanto, haver uma competição entre os indivíduos, resultando que, em cada geração, apenas uma pequena parte sobrevive.

Darwin, é claro, não conhecia a genética. E este foi sempre um ponto que não conseguiu conciliar com suas ideias, isto é, porque as características não tendem à uma maior uniformidade que à variabilidade. O *argumento de Jenkin*²⁵, aqui exposto de uma forma mais socialmente adequada por Dawkins (1985), era problemático para a teoria de Darwin. Sem o conhecimento da existência de genes, a evolução não poderia ser justificada.

We can rephrase Jenkin's argument in a more neutral analogy. If you mix white paint and black paint together, what you get is grey paint. If you mix grey paint and grey paint together, you can't reconstruct either the original white or the original black. Mixing paints is not so far from the pre-Mendelian vision of heredity, and even today popular culture frequently expresses heredity in terms of a mixing of 'bloods'. Jenkin's argument is an argument about swamping. As the generations go by, under the assumption of blending inheritance, variation is bound to become swamped. Greater and greater uniformity will prevail. Eventually there will be no variation left for natural selection to work upon (DAWKINS, 1985. p. 114.).

Fato 4: Os indivíduos em uma população apresentam grande variabilidade.

Fato 5: Muito dessa variação é hereditária.

Disto tudo, Mayr nos apresenta duas novas inferências:

Inferência 2: A sobrevivência não acontece aleatoriamente, mas depende, em parte, das características herdadas. O processo de seleção natural é fruto da desigualdade na sobrevivência.

²⁵ Proposto por Fleeming Jenkin em 1867, em inglês denominado *swamping argument*.

Inferência 3: Ao longo de gerações, a seleção natural acumula variações que levam a uma mudança gradual da população, eventualmente produzindo novas espécies.

Assim, listamos as principais implicações filosóficas da Teoria de Darwin, segundo Mayr:

A substituição de um mundo estático por um mundo em evolução, a demonstração da implausibilidade do criacionismo, a refutação da teleologia cósmica (a ideia de haver um propósito no universo), a abolição de qualquer justificação para um antropocentrismo absoluto (a ideia de que o propósito do mundo é a produção do humano), a explicação do 'design' no mundo por processos puramente materiais, a substituição do essencialismo pela lógica populacional (Mayr *apud* WATSON, 2006. Tradução nossa.)

No entanto, embora a teoria da evolução seja essencialmente correta, uma abordagem quantitativa compatível com as evidências arqueológicas ainda não foi obtida. E, neste mesmo sentido, Leslie Valiant comenta:

Quantitative theories of population dynamics have existed for a century. These theories are concerned with analyzing competition among static entities, and predicting how their relative population sizes will change as a result of competition. They have relevance to evolution, but do not address the question of how quickly organisms of increasing complexity can evolve. There is no theory known that explains quantitatively how competition by itself leads to greater functionality and complexity. Yet living things are highly complex mechanisms by any measure. On the part of Darwin, or his successors, of course, there is the excuse that they did not know—indeed, could not have known—what we now know regarding biochemistry and computation.(VALIANT, 2013).

Um caso que chama a atenção é a explosão de diversidade de espécies no Período Cambriano. O estudo dos registros fósseis revela que, antes do cambriano, os organismos eram relativamente simples, constituídos em grande parte por estruturas unicelulares ou pequenos organismos multicelulares. Porém, há aproximadamente 541 milhões de anos ocorreu uma rápida diversificação que deu origem à maioria dos filos hoje existentes(CONWAY-MORRIS, 2003, SNIR, YOHAY, 2019).

Estamos falando de um fenômeno tão complexo que mesmo as tecnologias mais avançadas de simulação, usando os mecanismos conhecidos de evolução, não conseguem replicar. Isso mostra como grande parte da equação que explicaria, cabalmente, a evolução continua incógnita.

3.3.1 SIMULAÇÃO DA EVOLUÇÃO BIOLÓGICA

Ao observar as sementes de salgueiro serem levadas pelo vento no fundo de seu jardim, Richard Dawkins (1985) comenta, “está chovendo DNA lá fora”.

It is raining instructions out there; it's raining programs; it's raining tree-growing, fluff-spreading, algorithms. That is not a metaphor, it is the plain truth. It couldn't be any plainer if it were raining floppy discs.(DAWKINS, 1985. p. 111.)

Usando como analogia os precários computadores dos anos 80, Dawkins explora o caráter digital do código genético e a origem da variabilidade dos organismos no acúmulo de sucessivos “erros” de cópia. Dawkins, através de seus organismos simulados, denominados *biomorfos*, está interessado em entender como a “evolutividade” evoluiu.

Imagine that a wide-open space of evolutionary opportunity has suddenly opened up - say a deserted continent has suddenly become available through natural catastrophe. What kinds of animals will fill the evolutionary vacuum? They will surely have to be descendants of individuals good at surviving in the post-catastrophe conditions. Interestingly, *some kinds of embryology might be especially good not just at surviving but at evolving.* (DAWKINS, 1985. p. 328.)(Destaque nosso).

Podemos pensar a evolução como um deslocamento em um espaço multidimensional, cada organismo que já existiu se encontra em algum lugar deste espaço, assim como todos os organismos que poderiam existir. Dawkins faz uma analogia, rerepresentando a ideia de um macaco frente à uma máquina de escrever, teclando aleatoriamente, produzindo – eventualmente – a obra completa de Shakespeare. Calcula a probabilidade de produzir uma única frase a princípio: “METHINKS IT IS LIKE A WEASEL”. A probabilidade de isto acontecer parece, a princípio, muito pequena para ser considerada, especificamente 1 em 27²⁸ tentativas. Mas se cada letra correta for mantida - ao invés de se descartar tudo e recomeçar – o processo rapidamente converge ao resultado esperado. Isto é, o acúmulo de pequenas modificações benéficas para o organismo não é puramente aleatório, mas a mutação é um trajeto guiado no espaço de possibilidades. Assim, a ideia da evolução da “evolutividade” pode ser entendida como um aperfeiçoamento da capacidade de navegar no espaço de possibilidades, através da história evolutiva. Estamos considerando o critério de seleção, ou *fitness*, o quanto o produto de cada geração se aproxima das obras de Shakespeare. Mas, para sermos justos,

poderíamos pensar em qualquer obra literária possível como um destino igualmente válido.

O primeiro cientista que usou computadores para se aprofundar nos mecanismos da seleção natural foi o virologista Nils Aall Barricelli (1912 – 1993). Ainda quando estas máquinas eram usadas para calcular a primeira bomba H, ele já experimentava com universos simulados no *Institute for Advanced Studies* (IAS). Dyson (2012) comenta que a motivação dele vinha de querer provar que a cooperação, e não a competição era o principal motor do processo evolutivo.

The author [Barricelli] has shown in several publications that evolution processes in many respects similar to biological evolution can now be obtained using self-reproducing entities of numerical nature whose properties are completely under control. (BARRICELLI, 1962)

Apenas um ano antes de Barricelli iniciar seus experimentos, Raymond Gosling havia feito a imagem por difração de raios-X, denominada *Photo 51*, que foi fundamental para Francis Crick e James D. Watson decifrarem a estrutura química do DNA: a natureza digital do código genético havia sido provada. Barricelli, porém, já sabia que a genética tinha que ser digital: Claude Shannon havia defendido a tese de PhD intitulada *An Algebra for Theoretic Genetics* em 1940 e von Neumann expressado que, para um organismo sobreviver em um ambiente imprevisível e dominado pelo ruído analógico, este precisaria usar meios digitais de correção de erro (DYSON, 2012).

A similaridade entre o funcionamento do DNA – ou, ao menos, a não total dissimilaridade – e uma máquina de Turing tem sido levantada desde que este modelo de computação se tornou conhecido. Outrossim, trabalhos mais recentes têm demonstrado os usos de computação natural para resolver problemas classificados como NP-completos (APÊNDICE A), por exemplo, o Problema do Caminho Hamiltoniano – um caso especial do Problema do Caixeiro Viajante – usando filamentos de DNA (ADLEMAN, 1994). Este método, inclusive, apresenta vantagens, se comparado à computação em silício, nos quesitos de permitir maior paralelismo e ter menor consumo energético (KARI, DALEY, *et al.*, 1999).

Uma iniciativa para esta discussão é a proposta de Gregory Chaitin, em seu livro *Proving Darwin, making biology mathematical*, no qual procura trazer “a precisão, generalidade e o grau de abstração da matemática pura” (CHAITIN, 2012).

Tradução nossa.) para as origens da biologia. Seu argumento é o de que a evolução por seleção natural, em sua essência, deve depender de um mecanismo simples e fundamental, que pode ser capturado – e simulado – por uma máquina de Turing.

O modelo teórico de Chaitin, denominado *Metabiology*, pretende explorar os elementos fundamentais para se descrever o fenômeno evolutivo, conservando apenas as partes que considera essenciais para dar origem ao mecanismo subjacente à evolução através da seleção natural. Concordamos com sua abordagem epistemológica no sentido de defender que o conhecimento sobre um processo só existe quando somos capazes de simular o mesmo.

As descobertas de Barricelli adicionam mais elementos à discussão: segundo o virologista, apontando uma limitação de suas simulações, após certo número de gerações seu mundo simulado tendia a ser dominado por um único organismo ao invés de gerar a crescente variedade compatível com o observado na natureza biológica.

E é nisto, que a proposta de Chaitin consegue ser, provavelmente, bem-sucedida, indefinidamente crescendo em complexidade. O critério de *fitness* utilizado é a geração de um número. Quanto maior o número, melhor. Mas não qualquer número: se pensarmos em um número binário, deseja-se um algoritmo capaz de imprimir a maior quantidade de “1”, sem entrar em um ciclo interminável – pois não haveria nenhuma dificuldade de se repetir uma fila interminável de 1. Em outras palavras, o objetivo é produzir um algoritmo do tipo *busy beaver* que, por sua vez, dá origem a um número com elevada complexidade. Assim, a complexidade evolutiva se conecta ao conceito, delineado por Ray Solomonoff, de complexidade algorítmica.

A mutação ocorre pela alteração de um bit no algoritmo. A maioria das alterações não levam a um programa mais funcional, ao contrário, a maior parte impede que seja realizada a função almejada: não produz um número como saída ou entra em *loop*.

Apesar de sua genialidade, o modelo de Chaitin tem objetivo diferente daquele que buscamos, não oferecendo contribuições suficientes em relação à pergunta central deste trabalho, que é sobre como um organismo pode aprender

sobre seu ambiente para maximizar seu sucesso. A abordagem da *Metabiology* não trabalha com população, e, portanto, nem com variabilidade entre indivíduos. Com um único agente, que não se reproduz, portanto, o fator hereditariedade não é incorporado no sistema. Com isso, a sobrevivência depende unicamente do critério de *fitness*.

Ademais, também não são considerados os recursos, e o que buscamos aqui é intrinsecamente ligado a limitação de recursos. Um organismo que demora demais para responder aos eventos do meio ou usa recursos excessivamente não está, a nosso ver, utilizando uma estratégia viável evolutivamente.

3.3.2 COMPUTAÇÃO EVOLUTIVA

Um processo largamente utilizado na indústria é a computação evolutiva, que se trata de um conjunto de meta-heurísticas que possui diversas formas de implementação, porém dentro de um molde que tenta simular aspectos de como se entende a evolução biológica.

O processo em geral utilizado pode ser visualizado na *Figura 8*, iniciando com uma geração aleatória de agentes que são avaliados com base em um critério de *fitness*. Os indivíduos que pontuaram melhor neste critério, são os selecionados para realização de troca de partes do código entre si, o que é denominado *crossover*, ou para sofrerem mutação. Estes novos indivíduos são avaliados contra os critérios de *fitness* e, os melhores, substituem na população aqueles que haviam antes pontuado menos.

O software *Avida*²⁶ é uma plataforma para o desenvolvimento de organismos digitais, disponível em Linux sob licença LGPL. Usando esta plataforma, em um cluster *Beowulf* de 64 máquinas *Pentium III*, Lenski et al. (2003) procuraram estudar como funções lógicas complexas poderiam emergir a partir em organismos digitais que possuíam apenas a capacidade de se autorreplicar.

²⁶ Disponível em <https://github.com/devosoft/avida>

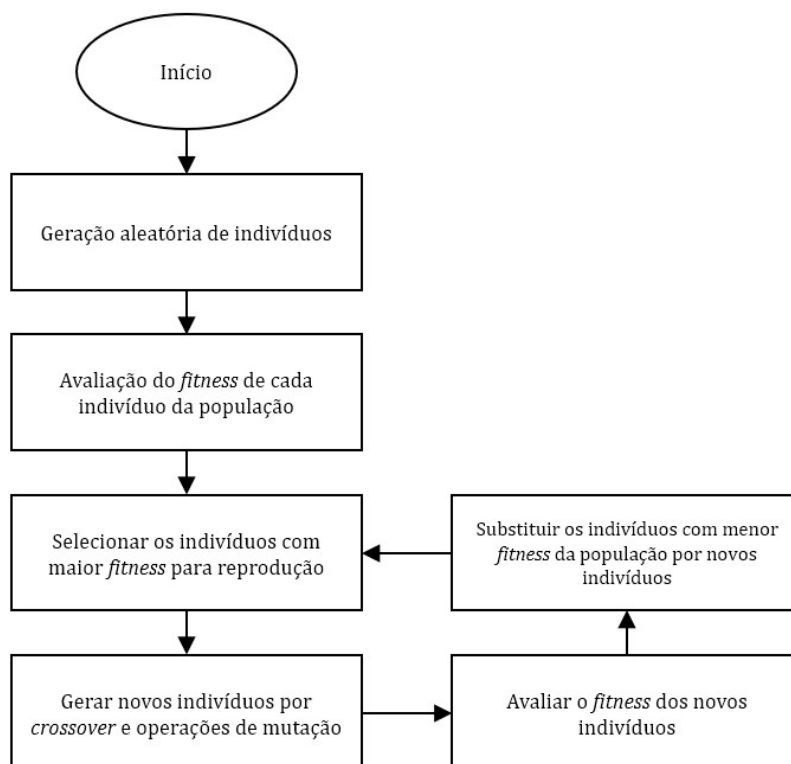


Figura 8: Esquema genérico de um fluxograma para computação evolutiva.

A plataforma proporciona um universo composto por 50 organismos que se reproduzem de forma assexuada. Cada organismo possui três registros, AX, BX, CX, de tal forma que a operação que avalia o *fitness*, conforme definido por Lenski et al (2003), compara os registros, tomando BX e CX como *input* e AX como *output*. Quando as operações nos bits de AX refletirem alguma função lógica, dentre uma lista (Quadro 9), o organismo será recompensado com mais ciclos de processamento. Apenas uma instrução do código genético realiza operação lógica, no caso, NAND²⁷. O objetivo é fazer emergir a operação EQU, a mais complexa da lista.

Os autores do estudo perceberam que a emergência da operação EQU aconteceu em uma sucessão de descobertas de operações mais simples. Isto é, o processo evolutivo construiu-se gradualmente a partir de mutações que produziram resultados úteis. Porém, quando as operações mais simples não eram recompensadas, isto impedia que funções complexas emergissem. Observou-se também que, no decorrer das simulações, algumas gerações manifestaram a perda

²⁷ Composição de “Negação” lógica e “E” lógico.

de funções já adquiridas, mas que isto foi sucedido pela invenção de funções ainda mais complexas. A perda parece ser um efeito colateral do processo que levou à produção de maior complexidade adiante. Isto, observam os autores, corrobora a posição de Darwin, quem considerava, há quase dois séculos, que a evolução de características complexas acontece pela modificação de estruturas e funções existentes. (LENSKI, OFRIA, *et al.*, 2003)

Quadro 9: Operações lógicas e Recompensas Fonte: Lenski et al (2003)

Function name	Logic operation	Computational merit
NOT	$\sim A; \sim B$	2
NAND	$\sim(A \text{ and } B)$	2
AND	A and B	4
OR_N	(A or $\sim B$); ($\sim A$ or B)	4
OR	A or B	8
AND_N	(A and $\sim B$); ($\sim A$ and B)	8
NOR	$\sim A$ and $\sim B$	16
XOR	(A and $\sim B$) or ($\sim A$ and B)	16
EQU	(A and B) or ($\sim A$ and $\sim B$)	32

3.3.3 MODELO PROVAVELMENTE APROXIMADAMENTE CORRETO

Probably Approximately Correct (PAC) é um esquema desenvolvido para a abordagem matemática das metodologias de *Machine Learning* (ML). Em uma visão sumária, trata-se de encontrar a melhor hipótese que explique um certo conjunto de observações. O conjunto de exemplos de certo evento usado em ML é sempre limitado, pois queremos que o algoritmo após o treino seja capaz de identificar novas instâncias não presentes no conjunto inicial. A partir destes exemplos esperamos produzir um algoritmo com certo grau de generalização. Por exemplo, em ser capaz de distinguir fotos de cachorros e gatos, o algoritmo precisa extrair características das fotos exemplo que permita identificar fotos que nunca teve acesso. Esta hipótese, portanto, pode no máximo ser *aproximadamente correta*, pois não é possível produzir um algoritmo a partir do conjunto de todas as fotos possíveis de cães e gatos: incorremos no problema da indução.

Aliás, é fato conhecido que cada conjunto de exemplos possui vieses, em geral desconhecidos por quem escolhe estes exemplos. O que leva o treino,

inadvertidamente, a produzir um algoritmo que atenta para padrões nos exemplos que podem não ser uma característica do evento que pretendemos generalizar. Mas são, não obstante, altamente eficientes na classificação daquele específico conjunto de exemplos. Este algoritmo, desta maneira, ao se deparar com novos casos, que não possuam o viés oculto, apresentará desempenho pobre. Isto é, tecnicamente, chamado *overfitting*.

Reconhecendo os limites da indução, o modelo PAC procura uma hipótese que generalize os exemplos de forma eficiente. Uma hipótese que trata cada caso isoladamente, como no caso de *overfitting*, não tem generalidade nem poder explicativo. E uma hipótese que não captura o fenômeno melhor que o acaso, também não é útil. Uma boa hipótese está entre estes dois extremos.

Valiant defende que a evolução é um algoritmo de aprendizagem, mais especificamente, um caso particular de aprendizagem PAC. Para distinguir os algoritmos que emergem nos processos naturais daqueles especificados por um programador, o autor cunhou o termo *ecorritmo*. Mas aqui, o fato de ser natural ou artificial, não é o aspecto relevante, e sim o fato de que quando um programador especifica o algoritmo ele possui uma teoria, correta ou não, de como as coisas são (*theoryful*). A evolução por seleção natural é *theoryless*: o ambiente não é totalmente conhecido e pode possuir muita arbitrariedade (VALIANT, 2013). Em muitos casos, inclusive, uma teoria pode sequer ser possível.

Os ecorritmos devem, no entanto, possuir elevada generalidade, pois, dadas as incertezas do mundo, um excesso de especificidade pode ser contra produtivo. Organismos vivos precisam se adaptar continuamente a um mundo que está sob as pressões das mudanças no espaço físico, como intemperismos, mas, também, da decorrente da adaptação dos outros organismos, como aparecimento de predadores ou extinção de uma fonte de nutrientes. Ademais, a resposta precisa acontecer em tempo hábil para as ações serem bem-sucedidas. Por exemplo, um dado tipo de organismo deve ser capaz de evadir seus predadores comuns com significativo grau de sucesso, sendo que estes predadores estão, geração após geração, em um processo de contínuo aprimoramento de sua capacidade de caça.

A aprendizagem, que é neste contexto a geração de ecorritmos, ocorre através da extração de regularidades do mundo. Estas regularidades são o conjunto de pressões evolutivas a que uma população está sujeita. Como organismos mais adaptados têm maior tendência a preservarem seus códigos genéticos através da hereditariedade, há uma seleção contínua de ecorritmos que 1) são menos custosos de se produzir e 2) são capazes de maior generalidade. Assim, a aprendizagem acontece ao se gerar uma hipótese sobre o mundo que é capaz de generalidade, mas conserve certo caráter sintético o suficiente para ser prática.

The requirements that such an algorithm [ecorithms] must meet to offer a plausible explanation of a natural phenomenon, such as biological evolution, are quite onerous. In particular, the algorithm must achieve its goals after a limited number of interactions and with the expenditure of limited resources. The concept of ecorithms and the general model of learning in which they are embedded, which I call probably approximately correct (or PAC) learning, insist on such quantitative practicality. The phenomena that they seek to explain are some of the most familiar to human experience: learning, resilience, and adaptation. I argue that broader phenomena still, in particular evolution and intelligence, are also best understood in these terms.(VALIANT, 2013)

Isto não é, senão, uma forma de enunciar a navalha de Occam: admitindo a constância de todos os fatores, a hipótese mais simples é a melhor. E isto é particularmente relevante, já que a evolução não possui uma teoria do mundo é preciso ter um mecanismo de selecionar quais as hipóteses têm mais probabilidade de corresponderem à generalização dos eventos.

Nem todos os eventos podem, no entanto, serem generalizados. Eventos aleatórios, por exemplo, não possuem uma hipótese com melhor generalização que a probabilidade de estar certa em 50% das vezes. É o caso de prever caras no lançamento de moedas. Alguns tipos de eventos requerem conhecimento prévio para serem corretamente generalizados: por exemplo, sem o conhecimento do DNA o funcionamento da hereditariedade permanecia. Por fim, há eventos que, apesar de não serem aleatórios, possuem uma complexidade que torna proibitiva a tentativa de generalização. Como exemplo, prever o clima no decorrer de um mês. Portanto, dentre os fenômenos que podem ser descritos através de algoritmos, apenas uma parte destes pode efetivamente ser decodificado, impondo um limite rígido para o que pode ser aprendido a respeito do mundo.

3.4 Conclusão e Discussão

Estamos nos encaminhando para admitir que a evolução, natural ou simulada, representa uma forma de se aprender sobre o mundo. Apesar de não conhecermos os detalhes de como a vida surgiu, não é contrassenso afirmar que um processo teve origem através da simples interação de matéria e energia, resultando em formas cada vez mais eficientes de se aprender. De fato, o código genético traz consigo uma enormidade de comportamentos necessários para sobrevivência dos organismos, mesmo daqueles que dependem fortemente de alguma forma de processo educacional desde o nascimento para serem bem-sucedidos, como é o caso de aves e mamíferos. E podemos, também, afirmar com segurança que, o aprendizado através da experiência, depende dos mecanismos acumulados através da história evolutiva. François Chollet, no Twitter²⁸, afirmou de uma forma que resume bem nossa percepção: “os *priors* cognitivos inatos que temos sobre o mundo a nossa volta levou milhões de anos para evoluir e, portanto, é remoto, e compartilhado com primatas não-humanos”. *Prior* é aqui utilizado para se referir à probabilidade *a priori* no teorema de Bayes. Continua Chollet, “Nossos *priors* sobre navegação espacial são ainda mais largamente compartilhados com espécies tão distantes quanto as aves” (Tradução nossa). Isto levanta a questão de quanto conhecimento precisa ser construído em um sistema de AI para melhor desempenhar sua função. E, principalmente, como este conhecimento é construído. Se estas extrapolações forem válidas, é um forte argumento contra o uso de abordagens de caráter simbolista, ontologias e sistemas especialistas – o que é, de fato, o que acontece atualmente na indústria.

O mundo natural é, irreconciliavelmente, irregular e, em si mesmo, desprovido de qualquer significado. No entanto, organismos vivos são capazes de encontrar suficiente estruturação, a tal ponto que são capazes de prever o estado do mundo no futuro. Quando um ser vivo, um inseto, por exemplo, nasce, ele já assume a presença de alimento, de predadores, do ciclo do dia e da noite, assim como de estações do ano, de parceiros sexuais, de várias formas e técnicas para obter

²⁸ Em 28 de novembro de 2020, às 14:23. O comentário se refere ao artigo de PAYNE, LYNCH, *et al.* (2020) sobre a representação espacial em pássaros. A discussão pode ser acompanhada através de <https://twitter.com/fchollet/status/1332736792605253632>.

alimento, enfim, uma vastidão de fatos e regularidades do mundo. Também, impõe um significado à realidade, na escolha dos elementos que são relevantes para sua tomada de decisão: apenas uma faixa do espectro está acessível à visão, uma pequena quantidade de ruídos é de interesse, apenas alguns tipos de formas serão objeto da atenção de seus cérebros, ignorando tudo o mais.

Crianças ao nascerem, por exemplo, parecem já possuir uma espécie de Física inata (HESPOS, VAN MARLE, 2012), isto só faz sentido se, além da irregularidade aparente, existir uma estrutura no mundo que pode ser extraída através do processo de aprendizagem. Provavelmente estes saberes inatos correspondem a uma grande quantidade, senão a maior parte, dos saberes necessários à nossa sobrevivência. E a adaptação das espécies ao ambiente nos quais se encontram representa esse contínuo movimento de transformação, através das eras, por meio do qual organismos acompanham a transformação do meio.

4 INTELIGÊNCIA EM MÁQUINAS: PODEM MÁQUINAS PENSAR?

4.1 Introdução

Quando se inicia o estudo de um assunto, um aspecto que costuma ganhar atenção é a nomenclatura dada à disciplina que se ocupa deste tema. Neste caso estamos falando da *Inteligência Artificial* (AI), termo cunhado por John McCarthy a partir da proposta de realização de um *workshop* em Dartmouth College, no ano de 1956. A escolha do nome foi com a intenção de distinguir este novo campo de estudo da Cibernética, cuja teoria de feedback analógico McCarthy discordava; e da Teoria de Autômatos, que era uma teoria matemática, na qual estava envolvido, mas que considerava por demais abstrata e estreita (NILSSON, 2010).

O termo “inteligência” é notoriamente difícil de ser definido. Similarmente, definir “artificial” (MONOD, 1971) não é uma tarefa particularmente fácil. Iremos nos deter, porém, apenas no termo “inteligência” tendo em vista que o fato de ser artificial ou não pode ser irrelevante no contexto da inteligência. E, além disso, “artificial” não é a parte que causa dificuldades (WANG, 2019).

O *Oxford Companion to the Mind*, uma obra de referência na área de Filosofia da Mente, traz algumas tentativas de definições de inteligência, coletadas em 1921 pelo *Journal of Educational Psychology*, de 17 especialistas. Depois de comentar algumas delas, Robert J. Sternberg, o autor do verbete, acena que, “estritamente falando, parece haver tantas definições de inteligência quanto houve especialistas convidados a defini-la” (GREGORY, ZANGWILL, 1987. Tradução nossa.). E quem estiver insatisfeito com a heterogeneidade das respostas pode tomar a definição circular de E. Boring (1924), ao afirmar que “a inteligência é o que os testes de inteligência testam” (BORING, 1961, GREGORY, ZANGWILL, 1987. Tradução nossa.). Em um sentido mais atual, a coleção mais completa de definições que pudemos encontrar foi feita por Legg e Hutter, em 2007. Nesta, separou as definições coletivas, listando 18 provenientes de dicionários e enciclopédias. As definições realizadas por profissionais do campo da psicologia corresponderam foram 35 e as definições propostas por pesquisadores de AI foram 18 (LEGG, HUTTER, 2007a).

Nosso objetivo é explorar o que se entende por inteligência na psicologia humana e no campo de cognição animal, para, posteriormente, nos debruçarmos sobre as tentativas de recriar a inteligência através de recursos computacionais. A partir daí, tentaremos extrair o que estas tentativas nos informam sobre o tema. O intuito é produzir uma visão geral que ofereça condições para o avanço de uma teoria geral da inteligência capaz de englobar tanto a biologia como a ciência da computação, para esclarecer nosso entendimento do funcionamento de organismos biológicos e, simultaneamente, acelerar o desenvolvimento tecnológico e a inovação. Dentro do conceito de inteligência, abordaremos desde as implicações de se entender inteligência restrita a um único domínio, como jogar xadrez ou identificar rostos humanos, até a ideia de uma inteligência mais geral, a qual seria capaz de usar eficientemente uma ampla variedade de estratégias em muitos ambientes diferentes.

Metodologicamente, nos amparamos em diversos autores de referência na psicologia para construir um mapeamento do termo inteligência. Partir deste âmbito tem especial relevância para um entendimento academicamente apropriado do conceito, uma vez que a psicologia é a área, por excelência, de estudo do tema. Estes autores e suas principais ideias são apresentados com algum detalhe no decorrer do texto. Para a abordagem do uso do termo inteligência aplicado a animais não-humanos foi feita uma pesquisa nas bases acadêmicas por artigos sobre cognição animal e inteligência animal, dentre os quais nos fundamentamos principalmente no trabalho de Dicke e Roth (2016), por fazer uma revisão sistemática de trabalhos sobre inteligência animal e por ser um dos raros que usou este termo. Paralelamente utilizamos pontualmente resultados de Herculano-Houzel (2012) e Herculano-Houzel, Mager et al. (2014). No que tange à inteligência de máquinas, utilizamos como base, principalmente, a tese de doutorado de Shane Legg, apresentada na Universidade de Lugano, em 2008, além da pesquisa de seu orientador, prof. Dr. Marcus Hutter, proponente do modelo AIXI.

4.2 Algumas dificuldades em se estudar a inteligência.

Como humanos, é desafiante evitar nossos próprios vieses ao definir inteligência. O que é compreensível por não termos outro exemplo – até o momento – de formas de inteligência mais exemplares que a humana para nos ajudar a

delimitar o conceito (BUCKNER, 2013, STRONGMAN, 2007). É mais fácil, como ponto de partida, abordar o comportamento percebido como inteligente do que discutir a noção de inteligência de forma abstrata.

Assim, se uma espaçonave alienígena aparecesse em nosso planeta, a noção de inteligência seria fortemente pressionada a incluir os criadores deste novo objeto²⁹. Poderíamos até discordar das decisões técnicas e estéticas que esses seres tomaram no projeto da espaçonave, mas dificilmente poderíamos imputar-lhes ausência de inteligência. Pois, a própria existência deste objeto manifesta a inteligência de seus criadores ao realizar um feito que não somos capazes mas desejaríamos produzir. Em contrapartida, ao observar o comportamento de animais não-humanos, o sentimento é bastante diferente, apesar de, em muitos casos, manifestarem habilidades que superam capacidades humanas, demonstrando alta complexidade cognitiva e sofisticação nas estratégias de resolução de problemas. No entanto, até mesmo o uso da palavra “inteligência” é muitas vezes evitado, preferindo-se falar em termos de habilidades cognitivas.

Frans de Waal, ao lado de uma série de exemplos em que animais apresentam habilidades técnicas e intelectuais comparáveis às humanas, reflete sobre a tradicional atitude antropocêntrica de nosso olhar e destaca a importância de entender-se os animais em seus próprios termos:

Experiments with animals have long been handicapped by our anthropocentric attitude: We often test them in ways that work fine with humans but not so well with other species. Scientists are now finally meeting animals on their own terms instead of treating them like furry (or feathery) humans, and this shift is fundamentally reshaping our understanding.(DE WAAL, 2013)

Embora a importância histórica da perspectiva cartesiana seja inegável, é fundamental reconhecer que ela nos legou uma concepção que deixa pouco espaço para vida interior dos animais. O dualismo cartesiano relegou os animais não-humanos à condição de meros autômatos. Sob este prisma, qualquer simpatia por eles seria desperdiçada em máquinas vazias, incapazes de qualquer senciência para receber nossa compaixão.

²⁹Cf. APÊNDICE B para uma discussão sobre a possibilidade de inteligência, mentes e consciência alienígenas extra-terrestres.

Uma certa miopia para o universo subjetivo do não-humano tem permeado o pensamento científico ao longo dos últimos 400 anos e, com isso, um vasto universo de experiências do mundo permaneceu encoberto por teorias que não faziam jus à complexidade que apenas estudos mais recentes têm desvelado.

O homem comum³⁰, como argumenta Searle, nunca duvidou da inteligência de seu cão (SEARLE, 1996), da astúcia de seu gato, entre tantos outros animais que compartilham os espaços em que vivemos. Adicionalmente, muitos contos e histórias orais retratam o complexo e vivo universo desses animais, apesar, é claro, da antropomorfização. O que implica certa forma ingênua, no senso comum, de, não apenas reconhecer inteligência no comportamento de outros animais, mas também certa confiança em nossa capacidade de hierarquizá-los por grau de inteligência. Esta mentalidade, que é, em certo sentido, parecida com a *scala naturæ*³¹ de Aristóteles – filósofo que, aliás, capturou muito da sabedoria do senso comum em seus textos –, compõe um rol no qual golfinhos e primatas encontram-se no topo, logo abaixo dos humanos, e os seres unicelulares, na base. Porém, se olharmos com atenção, esta lista, em muitas partes, poderia ser confundida com um rol de *afinidades*. O efeito de se negar a existência de mente aos animais que consumimos, por exemplo, é bem documentado: os porcos são percebidos como menos inteligentes que cães por lhes serem atribuídas finalidades diferentes, a daqueles é o consumo de sua carne e a destes é a companhia. Outro exemplo é o mugido das vacas, normalizado como ausente de qualquer intenção pelos trabalhadores acostumados ao manejo (BASTIAN, LOUGHNAN, *et al.*, 2012, BRATANOVA, LOUGHNAN, *et al.*, 2011, GRANDIN, 2006, LOUGHNAN, HASLAM, *et al.*, 2010).

Ao submetermos animais a testes de inteligência, mesmo sob as condições mais cuidadosas, se percebe uma distribuição desigual de competências cognitivas

³⁰ Esta expressão, “o cidadão comum”, remete às expressões francesas *homme de la rue*, *le citoyen moyen* ou *citoyen λ*, como usado por John Searle. Apesar de discordarmos da forma com que muitas vezes Searle argumenta, pois é mais carregado de retórica que de argumentação filosófica, ele captura bem este descompasso entre a Filosofia – e, por consequência, a Ciência – e a percepção do senso comum. Este último, muitas vezes, como aqui, mais sagaz e fundamentado.

³¹ No texto *Da História dos Animais*, escrito no Séc. IV AEC, Aristóteles expressa sua visão da biologia na qual todos os seres estão dispostos em uma escala fixa de perfeição que ficou conhecida como a *scala naturae* ou *cadeia dos seres*. Seu sistema compunha-se pela gradação segundo as potencialidades de cada ser, desde os minerais até o humano no topo.

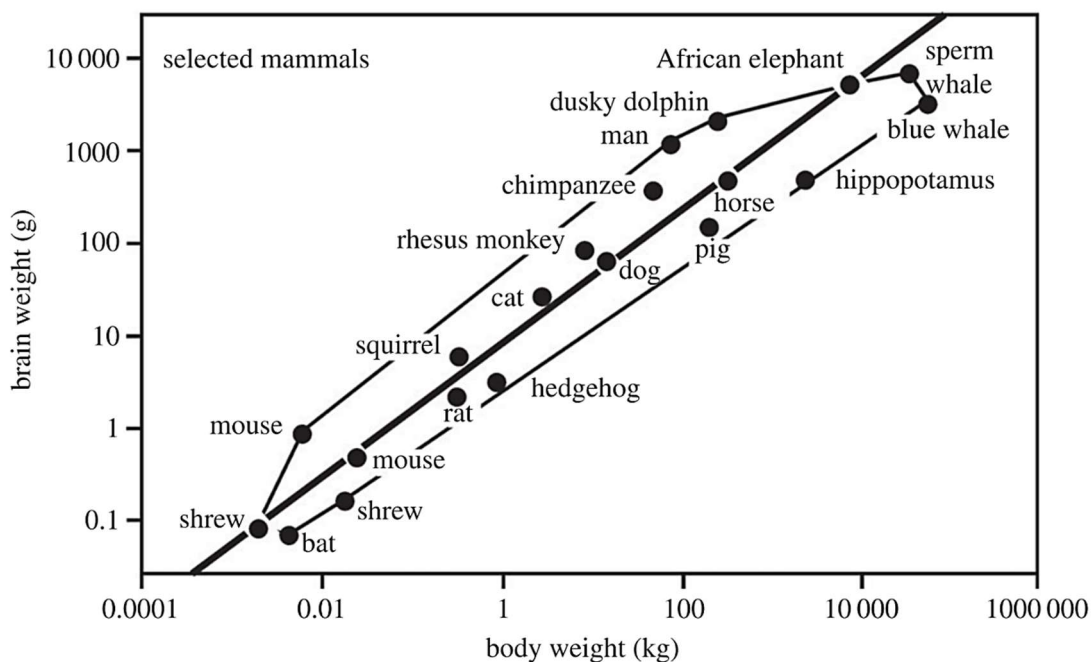
particulares. Sabemos que cães possuem uma capacidade, selecionada artificialmente, para entender nossas intenções, como o olhar e o gesto de apontar, enquanto seus parentes mais próximos, os lobos, não apresentam a mesma facilidade. E, com isto, os cães podem nos parecer – ou transparecer – dotados de grande inteligência. No entanto, os primatas em geral, que são usualmente apontados por especialistas como mais inteligentes³² que canídeos (DICKE, ROTH, 2016, MACLEAN, HERRMANN, *et al.*, 2017), têm problemas para seguir pistas visuais de nossa comunicação (HOPKINS, RUSSELL, *et al.*, 2013). Dificuldade similar se apresenta na tentativa de comparar as capacidade cognitivas de cães e porcos (LAZZARONI, MARSHALL-PESCINI, *et al.*, 2020, PÉREZ FRAGA, GERENCSÉR, *et al.*, 2020).

Quem tenha observado cães e gatos domésticos resolverem problemas no seu dia a dia se dá conta de que cada um aborda um mesmo problema de forma muito distinta do outro (MIKLÓSI, PONGRÁCZ, *et al.*, 2005). Acreditamos que estas diferenças são indícios de que a cognição, moldada por diferentes pressões evolutivas, dá origem a distintas especializações (BRÄUER, HANUS, *et al.*, 2020), ao invés de somarem-se para constituir uma única inteligência geral. Pela mesma lógica, portanto, a classificação de espécies em uma hierarquia por graus de inteligência nos parece inadequada. Mais coerente parecem ser as correntes que procedem a fragmentação daquilo que é percebido como comportamento animal inteligente em habilidades cognitivas distintas.

³² Para estas comparações, traduzindo Roth, define-se inteligência como “a flexibilidade mental ou comportamental de um organismo para resolver problemas que ocorrem no ambiente social e natural, culminando na emergência de soluções inovadoras que não são parte do repertório normal do animal” (DICKE, ROTH, 2016).

Se ao invés de comparar as competências de cada espécie para tentar inferir o grau de inteligência, tentarmos inferi-lo através da comparação entre características físicas dos cérebros das diferentes espécies, notaremos que também esta abordagem oferece pouco poder preditivo.

Figura 8: Relação entre massa do cérebro e massa corporal entre mamíferos.



Fonte: DICKE; ROTH, 2016

Na *Figura 8*, a linha de regressão, representada em escala logarítmica dupla, estabelece a média para mamíferos. Assim, os animais que aparecem acima devem ser significativamente mais inteligentes do que aqueles que se encontram sobre ou abaixo da linha, para uma mesma massa corporal. Entretanto, o estudo de Ursula Dicke e Gehrard Roth nota que, embora os humanos se sobressaiam mais do que qualquer outra espécie, há demasiada proximidade entre humanos, golfinhos e outros primatas. A partir disto, os autores concluem que o tamanho relativo do cérebro, isoladamente, não apresenta uma boa correlação com as diferenças de inteligência observadas entre estas espécies (DICKE, ROTH, 2016).

É necessário levar em consideração muito mais variáveis, como a quantidade de neurônios corticais e a densidade neuronal, distância entre neurônios e condutividade dos axônios (HERCULANO-HOUZEL, S., 2012, HERCULANO-HOUZEL, Suzana, MANGER, *et al.*, 2014). E, mesmo assim, corvídeos e psitacídeos permaneceriam requerendo explicações *ad hoc* para o alto grau de inteligência

percebido apesar do cérebro relativamente pequeno (DICKE, ROTH, 2016). Sem mencionar detalhes do extraordinário caso do sistema nervoso de cefalópodes, que o filósofo Peter Godfrey-Smith julgou os seres mais próximos de uma inteligência alienígena (GODFREY-SMITH, 2013).

Retomando o tema da AI, ficou evidente nos anos que se seguiram ao *workshop* em Dartmouth College a existência de uma fragilidade em nossa capacidade de reconhecer a complexidade dos atos de nossa própria inteligência. Para surpresa dos pesquisadores, resolver problemas de Cálculo e provar teoremas do *Principia Mathematica* é bem mais simples que falar ou descrever objetos do mundo apreendidos através do sentido da visão. Andar em duas ou quatro patas se mostrou um problema formidável com o qual jogar xadrez sequer se compara (MCCORDUCK, 2004, NEWELL, SIMON, 1956, ROSENBLATT, 1958, RUSSELL, NORVIG, 2020).

Hoje, em retrospecto, parece um tanto óbvio que os desafios da AI são imensos. A irregularidade do mundo natural demanda um poder computacional e uma complexidade algorítmica que surpreendeu algumas das mentes mais brilhantes que já viveram, incluindo Alan Turing (OLAZARAN, 1996, TURING, 1950). Engenheiros de software chamam de *brittleness* a dificuldade de sistemas de AI acomodarem elementos que estão além de um repertório fixo de situações predefinidas. Organismos biológicos, por sua vez, são formidáveis em se adaptar a pequenas variações inesperadas do ambiente: bactérias se adaptam a cada novo antibiótico, pombos passaram a viver e prosperar em centros urbanos, polvos aprenderam a invadir barcos pesqueiros para capturar os caranguejos já mortos. Enquanto isso, softwares de AI ainda podem ter dificuldade de entender que um saco plástico trazido pelo vento em direção ao para-brisa não é razão para subitamente frear o automóvel (DOMINGOS, 2015).

A conhecida frase, abaixo citada, do New York Times de 1958, que prometia fazer-se emergir dos dispositivos computacionais características cognitivas típicas de seres biológicos e, particularmente, humanos, foi profundamente danosa para reputação da área de AI. Promessas excessivamente otimistas acabaram colaborando para a retração de investimentos e, conseqüentemente, da pesquisa na área. Mais de sessenta anos depois, essa promessa continua elusiva.

WASHINGTON, July 7 (UPI) -- The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.(THE NEW YORK TIMES, 1958)

Marvin Minsky (1927 - 2016) comenta, em seu livro *Society of Mind*, publicado originalmente em 1986, que temos pouca consciência das capacidades que nossas mentes desempenham bem. Temos mais consciência de processos simples que não funcionam bem do que de processos complexos que funcionam perfeitamente (MINSKY, 2007). Isso se dá devido ao fato de a evolução biológica através de bilhões de anos de aperfeiçoamentos ter tornado inconscientes um grande repertório de percepções e comportamentos sensório-motores. Para Moravec a razão é uma tênue camada do pensamento humano que ainda não atingiu suficiente destreza e, por isso, nos parece difícil de exercer (MORAVEC, 1988).

4.3 A Inteligência Humana

Até o momento discutimos que a inteligência está parcialmente correlacionada à constituição do cérebro de uma forma bastante complexa. Por isso, nenhuma metodologia, até o momento, nos garante estimar a inteligência a partir do mero estudo do substrato neural. Também apontamos que são necessários muito mais recursos cognitivos para se realizar os movimentos do corpo e processar os sentidos do que para a abstração e o pensamento racional. E, acima de tudo, indicamos que nosso olhar é enviesado, sendo movido por ideologias e hábitos quando julgamos a inteligência de um outro ser, humano ou não-humano. Aliás, este é um tópico sensível, a respeito do qual a história é profusa de exemplos, como bem relata Stephen Jay Gould (1941 - 2002) em *The Mismeasure of Man*. Esta obra relata como o viés de confirmação, guiado pelo preconceito, levou eminentes cientistas a considerar mulheres, negros e pobres como intelectualmente inferiores, justificando a subserviência destes como um fato da natureza. Como exemplo, citamos os resultados relatados pelo renomado anatomista e antropólogo Paul Broca (1824 - 1880):

In general, the brain is larger in mature adults than in the elderly, in men than in women, in eminent men than in men of mediocre talent, in superior races than in inferior races (p.304). Other things equal, there is a remarkable relationship between the development of intelligence and the volume of the brain (p. 188). (*Apud* GOULD, 1992, p.102)

4.3.1 DEFINIÇÕES DE INTELIGÊNCIA HUMANA

Partindo da coleção citada anteriormente, feita por Legg e Hutter, de definições de inteligência dadas por profissionais da Psicologia, extraímos certos temas recorrentes, dentre os quais, sem dúvida, a adaptação ao ambiente é o que mais se sobressai. Das 35 definições, 12 mencionaram a adaptação e em 10 delas, esta foi a característica principal da definição. Neste tema, S. S. Colvin coloca que “uma pessoa possui inteligência proporcional ao quanto aprendeu, ou pode aprender, a se ajustar ao ambiente” (*apud* LEGG, HUTTER, 2007. Tradução nossa.). Em seguida, os temas mais comuns foram solução de problemas ou performance de comportamentos direcionados a um objetivo, com nove definições e, em sete delas, estes foram os temas centrais. Como exemplo, W. V. Bingham afirma que “devemos usar o termo ‘inteligência’ para significar a habilidade de um organismo de resolver problemas novos” (*apud* LEGG, HUTTER, 2007. Tradução nossa.). Pensamento abstrato foi a terceira característica mais presente na coleção, com oito definições, das quais seis consideram esta característica essencial. Neste sentido, L. M. Terman expressou que “[Inteligência é] a capacidade de realizar pensamento abstrato” (*apud* LEGG, HUTTER, 2007. Tradução nossa.). As demais definições foram menos específicas, algumas inclusive expressando o quanto a inteligência pode ser elusiva. “O termo inteligência designa um conjunto de funções complexamente inter-relacionadas, das quais nenhuma é completamente ou precisamente conhecida do homem (...)” (R. M. Yerkes and A. W. Yerkes *apud* LEGG, HUTTER, 2007. Tradução nossa.).

Dessa coleção, quatro definições subscrevem à perspectiva de que a inteligência é uma característica unificada, tal como expresso por Jensen em “inteligência é um fator geral que se expressa em todo o tipo de performance” (*apud* LEGG, HUTTER, 2007). E quatro acenam para a ideia de inteligência como uma reunião de capacidades:

Intelligence is not a single, unitary ability, but rather a composite of several functions. The term denotes that combination of abilities required for survival and advancement within a particular culture. (A. Anastasi *apud* LEGG, HUTTER, 2007)

4.3.2 MENSURAÇÃO DA INTELIGÊNCIA HUMANA

Apesar da dificuldade de se definir a inteligência e mesmo do debate ético no entorno da questão (WARNE, 2019), na psicologia os testes neste âmbito são razoavelmente bem sucedidos (KAUFMAN, RAIFORD, *et al.*, 2016, OPEN SCIENCE COLLABORATION, 2014). Testes de inteligência, normalmente denominados testes de IQ, possuem elevada reprodutibilidade e seus resultados são estáveis no decorrer da vida do indivíduo, além de possuírem significativa capacidade de prever o seu sucesso acadêmico. (TERMAN, MERRILL, 1960, Terman, Oden, 1959).

O Estatístico Francis Galton (1822-1911) foi pioneiro nos testes de inteligência. Seus testes focaram em tempo de reação, coordenação física e capacidade de discriminar sons: sua hipótese era a de que a velocidade de reação a um estímulo estava positivamente correlacionada com o grau de inteligência. Seus testes, porém, não apresentavam a precisão necessária para justificar sua hipótese, nem os métodos estatísticos disponíveis na época permitiriam derivar tal conclusão. Todavia, em 1985 seus dados foram revisados e revelaram que Galton estava, de fato, apontando na direção correta (JOHNSON, MCCLEARN, *et al.*, 1985).

Em seu livro de 1892, *Hereditary Genius*, Galton percebeu que na linhagem familiar de pessoas renomadas era frequente mais de um parente apresentar notável inteligência, além disso, esses indivíduos podiam se destacar em áreas diferentes. Desta observação, presumiu que a inteligência como habilidade geral é mais hereditária que as habilidades específicas, como o talento para arte ou para a matemática, por exemplo, sendo estas mais influenciadas por questões ambientais (JENSEN, 2002).

Alfred Binet (1857 – 1911) desenvolveu para o governo francês, em 1905, o primeiro teste na forma de pequenos problemas relacionados a tarefas do cotidiano, em espírito, similar aos testes utilizados atualmente. Os problemas envolviam comparações de pesos e dimensões de objetos, contagem de moedas, memorização de números e palavras, entre outros, organizados em ordem de dificuldade. O resultado era obtido através da normalização da pontuação do indivíduo para crianças na mesma idade. O objetivo era identificar crianças com problemas de aprendizagem. O teste de Binet foi adaptado por Lewis Terman (1877-1956) para

crianças estadunidenses, tendo passado por vários aprimoramentos, o que deu origem ao *Teste Stanford-Binet* (1916).

Charles Edward Spearman (1863 – 1945), influenciado por Galton, sugeriu que os resultados de diferentes testes de inteligência estariam correlacionados por um único fator de inteligência geral, que denominou fator *g*. Spearman enfatiza, em um forte veio positivista, que *g* é um fato estatístico que se distingue da habilidade apresentada pelo indivíduo em realizar uma única tarefa específica “sob certas condições o score de uma pessoa em um teste mental pode ser dividido em dois fatores, um que é o mesmo em todos os testes, e outro que varia de teste para teste” (Spearman *apud* DEARY, LAWN, et al., 2008. Tradução nossa).

Spearman também compara *g* a uma forma de energia, curiosamente – embora típico de sua época –, descrevendo-a como uma propriedade da física:

On weighing the evidence, many of us used to say that this *g* appears to measure some form of mental energy. But in the first place, such a suggestion is apt to invite needless controversy. This can be avoided by saying more cautiously that *g* behaves as if it measured an energy. In the second place, however, there seems to be good reason for changing the concept of energy to that of *power* (which, of course, is energy or work divided by time). In this way, one can talk about mind power in much the same manner as about horsepower (...) (Spearman *apud* DEARY, LAWN, et al., 2008)

Conclui que cada pessoa possui uma certa quantidade desse *poder* e este não pode ser treinado ou ampliado, da mesma maneira que, nos termos de Spearman, “uma pessoa não pode ser educada para ser mais alta” (Spearman *apud* DEARY, LAWN, et al., 2008. Tradução nossa.).

David Wechsler (1896-1981), que foi orientando de Spearman, modificou os testes para reduzir a ênfase nas habilidades verbais e, assim, incluiu problemas com imagens, tais como *puzzles* de figuras. Em 1939 desenvolveu o *Wechsler Adult Intelligence Scale* (WAIS), em 1949 o *Wechsler Intelligence Scale for Children* (WISC) e em 1967 o *Wechsler Preschool and Primary Scale of Intelligence* (WPPSI). Com isto, ao invés de um único número para o score de inteligência, passou-se a utilizar duas áreas, *verbal* e *performance*.

Robert Sternberg propôs três formas de inteligência, chamada *Triarchic Theory of Intelligence*, conforme *Quadro 10*:

Quadro 10: As três formas de inteligência definidas por Sternberg

Composicional	Capacidade de dividir problemas em partes e encontrar soluções não óbvias.
Experiencial	Refere-se à capacidade de se adaptar a novas situações e, após isso, automatizar a tarefa. É a capacidade de síntese.
Prática	É a capacidade do indivíduo de se adaptar às situações que o ambiente apresenta, integrando as duas formas anteriores de inteligência. Envolve competências de adaptação ao ambiente, transformação do ambiente e seleção do ambiente conforme melhor se ajustem aos objetivos do indivíduo.

Louis Leon Thurstone (1887-1955) considerou que a inteligência seria composta por um conjunto de *habilidades mentais primárias*, opondo-se à noção de uma única inteligência geral. Sua teoria de fatores múltiplos ressalta sete habilidades não relacionadas: compreensão verbal, fluência vocabular, facilidade com números, visualização espacial, memória associativa, velocidade de percepção e raciocínio. Sua contribuição influenciou, através das diversas revisões, no desenvolvimento dos testes de inteligência como *Stanford-Binet* e *WAIS*, descartando a noção de idade mental em favor de modelos de inteligência hierárquicos e análise fatorial (LEGG, 2008, MULAİK, 2010).

Em 1963, Raymond Cattell (1905 – 1998), amparado pela análise fatorial e pelo trabalho de Spearman, propôs os conceitos de inteligências *fluida* (g_f) e *cristalizada* (g_c). A inteligência fluida refere-se à capacidade de resolver novos problemas com pouca necessidade de conhecimento prévio, enquanto a inteligência cristalizada está ligada ao acúmulo de experiências que auxiliam em atingir um objetivo. Assim, g_f e g_c são fatores de g . A inteligência fluida tende a desenvolver-se com a idade, até a adolescência, a partir de quando começa a, gradualmente, reduzir. E a inteligência cristalizada continua a crescer durante a vida do indivíduo. Esta teoria foi sintetizada na *Cattell-Horn-Carroll theory*, estabelecendo-se como um dos modelos de inteligência mais bem aceitos na psicologia humana e, portanto, uma das

principais referências por detrás da formulação dos atuais testes de IQ (HORN, 1968).

Em contrapartida, algumas teorias, ao invés de buscarem unificação em um só fator, procuraram a via oposta, identificando dimensões irreduzíveis da inteligência. A teoria mais notável é a de Joy Paul Guilford (1897 – 1987), que rejeita a perspectiva de reduzir a inteligência a um único número, tal como proposto por Spearman. Segundo sua teoria da *Estrutura do Intelecto*, de 1955, a performance em testes de inteligência depende de 180 fatores, em três categorias: operações, conteúdo e produtos (*Figura 9*).

O modelo de Guilford foi criticado pela dificuldade de se mensurar tantos elementos, tornando inviável a realização de testes de inteligência na prática. Não obstante, muito de seu trabalho foi incorporado nos testes aplicados nas forças armadas estadunidenses.

De modo semelhante, Howard Gardner propôs, em seu livro *Inteligências Múltiplas*, de 1983, a existência de oito habilidades que se manifestam como inteligências: musical, visual-espacial, verbal-linguística, lógico-matemática, corporal-cinestésica, interpessoal, intrapessoal e naturalista. Entretanto, deixou em aberto para que novas inteligências fossem futuramente adicionadas. Ao contrário das teorias anteriormente descritas, a de Gardner se opõe à prática da mensuração de inteligência tal como é realizada pelos testes de IQ. Apesar do sucesso popular de sua teoria, não há evidências que a suportem, vinculando-a mais ao público leigo do que à pesquisa acadêmica (DINHAM, 2014, LEGG, 2008, WATERHOUSE, 2006).

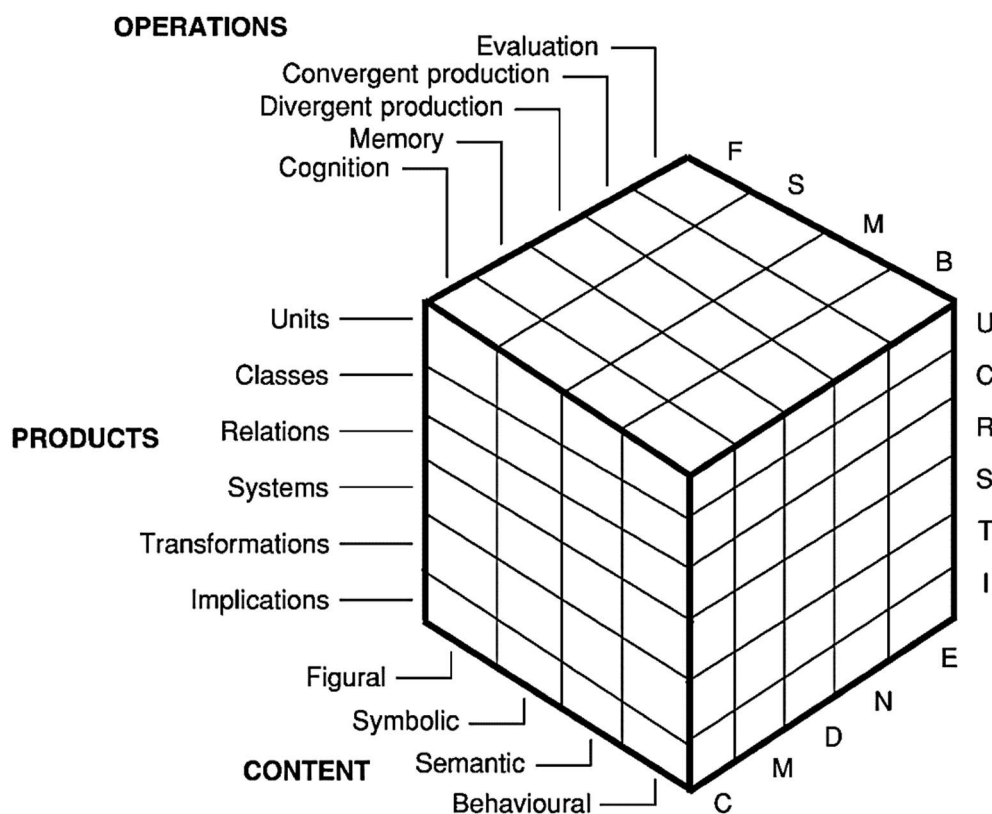


Figura 9: Representação da *Estrutura do Intelecto* como relação de três vetores unidimensionais formando um cubo. Fonte: MOSELEY, D., BAUMFIELD, V., ELLIOTT, J., et al. *Frameworks for thinking a handbook for teaching and learning*. 1st. ed. Cambridge, Cambridge University Press, 2005.

Este debate sobre inteligência tem, em seu interior, um conflito a respeito da concepção sobre a natureza humana: de um lado temos a perspectiva da inteligência como um conjunto de ferramentas específicas, temperado através da evolução biológica, compondo um amálgama de competências. Esta perspectiva engessa as habilidades dentro de moldes especificados por meio da pressão seletiva através de eras. E, ainda, propõe uma inteligência estática, gravada na estrutura dos cérebros desde o aparecimento da espécie. Por outro lado, existe a perspectiva da *tábula rasa* (PINKER, 1999), em que a inteligência é uma capacidade de se transformar experiências em conhecimentos e habilidades, para a qual nenhum problema está aquém de suas investidas. Esta perspectiva, por sua vez, descende da razão cartesiana e se alinha ao pensamento religioso de matriz judaico-cristã, ao antropocentrismo e, conseqüentemente, ao especismo (BRÄUER, HANUS, *et al.*, 2020).

4.4 *Inteligência em Inteligência Artificial*

O primeiro texto de referência sobre a inteligência em máquinas é o de Alan Turing, que publicou um artigo em *Mind*, um periódico de Filosofia, intitulado *Computing Machinery and Intelligence*, no ano de 1950. Neste texto fundador, Turing prescreve um teste segundo o qual uma máquina pode ser propriamente classificada como inteligente. Ele parte da hipótese de que um computador digital que venha a existir em qualquer futuro, independente da tecnologia utilizada, é equivalente a uma máquina do presente³³. Isso se aplicaria tanto em 1950, quando o artigo foi escrito, quanto nos dias atuais, pois todas as suas operações poderiam ser reduzidas às de uma máquina de Turing. Partindo desse pressuposto, seu artigo questiona se uma máquina imaginável poderia exibir uma inteligência similar à de um ser humano, a ponto de ser de fato confundida com um.

Perguntar-se sobre a possibilidade de uma máquina pensar é, de toda forma, uma pergunta vazia, estabelece Turing, pois requer uma definição do que se entende por pensar e do que se considera uma máquina. Seu primeiro interesse é livrar-se de certa parcialidade ou favoritismo antropocêntrico, pois em geral, consideramos *pensar* algo que uma pessoa faz. Porém, o conceito de pensar se torna gradualmente deslocado ao aplicá-lo a outros seres vivos, e, muito mais, se aplicado a um produto da engenharia, reflete. E, se levado ao extremo, não podemos sequer saber se outra pessoa pensa realmente, uma vez que apenas os nossos estados mentais nos são dados a conhecer. Quanto aos demais, apenas podemos fazer uso da analogia para inferir que seus estados mentais também existem. Turing propõe que a própria palavra “pensar” não representa adequadamente nossa dúvida, e acredita que ao final do século XX, dizer que um computador está pensando não causará imediato desconforto. Prevendo isto, substitui a pergunta original por uma mais bem formulada: seria uma máquina capaz de exibir inteligência similar à humana?

O procedimento que ele propõe, para que seja avaliada a capacidade de uma máquina, veio a ser conhecido como Teste de Turing, e consiste, basicamente, em implementar um algoritmo de imitação de ser humano que, quando capaz de

³³ Turing cita a Máquina Analítica a título de comparação, ressaltando que apesar de mais lenta, meramente mecânica e movida à vapor, ela seria totalmente compatível com as operações de um computador contemporâneo.

convencer um juiz de se tratar de uma pessoa humana, em ao menos 30% das tentativas³⁴, após cinco minutos de interação por escrito, seria considerado aprovado. Ou seja, seria concedido que esta máquina tem inteligência similar à de um ser humano.

Assim como uma *caixa preta*, as operações que tal máquina realizaria para produzir a imitação não devem ser alvo de objeções, posto que só nos interessa o efeito observável, a saber, a capacidade de se passar por uma pessoa humana, que é avaliada por meio da análise de sentenças por ela produzidas (*output*) em resposta às perguntas livremente escolhidas pelos juízes (*input*).

Ademais, se trata de reconhecer similaridade qualitativa, não quantitativa. Pois se poderia argumentar que tal máquina, na verdade, possuiria inteligência superior a um humano, posto que ela precisaria, intencionalmente, cometer erros e produzir demora em apresentar resultados de operações matemáticas para não se revelar ao juiz. Isto é, replicar as funções cognitivas humanas é, apenas, um dos atributos daquela máquina. Por exemplo,

Q: Please write me a sonnet on the subject of the Forth Bridge.

A: Count me out on this one. I never could write poetry.

Q: Add 34957 to 70764

A: (Pause about 30 seconds and then give as answer) 105621.

Q: Do you play chess?

A: Yes.

Q: I have K at my K1, and no other pieces. You have only K at K6 and R at R1. It is your move. What do you play?

A: (After a pause of 15 seconds) R-R8 mate. (TURING, 1950)

Não obstante, resta ainda definir o que se entende por máquina. Ciente de que o termo é demasiado amplo e que um ser humano poderia ser apresentado como exemplar de máquina, decide Turing, ao invés de delimitar as características que definem o que é uma máquina, o que considera uma tarefa ineficiente, arbitrariamente escolhe apenas debater, exclusivamente, o caso de computadores digitais. Justifica esta escolha por terem sido estes os dispositivos que levantaram, primeiramente, as questões ora debatidas. Além do mais, Turing comenta que “esta restrição parece ser drástica à primeira vista (...), mas será apenas insatisfatória se

³⁴ Por estranho que pareça não ter escolhido uma proporção de 50%, Turing deu esse número estimando o estado da arte no ano 2000 e, assim, a proporção 30% já seria indício suficiente para responder sua questão “podem máquinas pensar?” com um qualitativo “sim”. Pela via de argumentação que utilizou é possível inferir que acreditava que o progresso tecnológico eventualmente iria cobrir a diferença.

(contrariamente à minha crença), computadores digitais forem incapazes de se saírem bem no jogo [de imitar].” (TURING, 1950. Tradução nossa.)

Fica evidente que Turing, posto que considera o ser humano uma espécie de máquina, volta a sua indagação para a possibilidade de implementar em um computador digital um conjunto de instruções que seja suficiente para reproduzir os atributos, que geralmente identificamos com o comportamento intelectual, de um ser humano.

Em 1958 foi publicado, postumamente, um livro de John von Neumann (1903 – 1957) intitulado *The Computer and the Brain*. Neste, sustentou que o cérebro humano realiza um trabalho similar ao de um computador, isto é, cérebros são órgãos que realizam processamento sobre as informações sensoriais. Sua perspectiva é a de que uma teoria poderia unificar computadores digitais, cérebros e o DNA em uma mesma descrição. Neste sentido, compara o funcionamento do sistema nervoso a operadores lógicos, similar ao implementado por meio das válvulas ou transistores nos circuitos digitais:

I observed before that this particular mechanism—the stimulation of nerve pulses by suitable combinations of other nerve pulses—makes the neuron comparable to the typical basic, digital, active organ. To elaborate this further: if a neuron is contacted (by way of their synapses) by the axons of two other neurons, and if its minimum stimulation requirement (in order to evoke a response pulse) is that of two (simultaneous) incoming pulses, then this neuron is in fact an “and” organ: it performs the logical operation of conjunction (verbalized by “and”), since it responds only when both its stimulators are (simultaneously) active. If, on the other hand, the minimum requirement is merely the arrival (at least) of one pulse, the neuron is an “or” organ—i.e. it performs the logical operation of disjunction (verbalized by “or”), since it responds when either of its two stimulators is active. (NEUMANN, KURZWEIL, *et al.*, 2012)

Em seu texto original não usa a palavra inteligência uma única vez, embora seus prefaciadores, Ray Kurzweil e Paul e Patrícia Churchland usem com frequência. O que é compreensível, pois dá claros indícios de que as capacidades do cérebro podem ser, de alguma maneira, eventualmente traduzidas para um computador digital artificial.

Jeff Hawkins publicou em 2004 um livro com o título *On Intelligence*. Sua perspectiva é nascida da sua experiência como engenheiro elétrico e inventor: pretende entender os mecanismos através dos quais o cérebro dá origem à inteligência para criar o mesmo tipo de processo com os recursos computacionais.

Uma de suas mais interessantes observações é a de que, para ele, o cérebro não é primariamente um processador de informações, mas um preditor: compara os estímulos de experiências atuais a outras instâncias similares codificadas na memória, que chama de *representações invariantes*³⁵, com isto predizendo eventos que tendem a acontecer em seguida. Comportamento e predição são a mesma coisa, pois ao se realizar uma ação, o cérebro imagina a ação sendo realizada e os neurônios motores disparam, produzindo o comportamento. Somente quando a predição falha, é que a atenção é direcionada para, então, realizar o processamento da anomalia.

O aparato da inteligência, acredita Hawkins, é a organização hierárquica e recursiva do sistema nervoso, que funciona em via de mão dupla. A percepção sensorial, desta forma, acontece através da relação entre os padrões de ativação oriundos dos sentidos e a análise de níveis superiores, que continuamente direcionam e filtram o que é percebido. Esta estrutura se replica, nível após nível, cada vez com um grau maior de abstração até, finalmente, poder se tornar objeto da experiência consciente.

Seu argumento se direciona para explicar a diversidade de funções sensoriais não por diferentes mecanismos, mas a partir da organização dos componentes do sistema nervoso. Acredita, assim, que o neocórtex implementa o mesmo algoritmo básico em toda sua extensão, o que fundamenta na sua leitura do neurofisiologista Vernon Mountcastle (1918 – 2015).

From the signature example you can see that invariant representation in motor cortex is, in some ways, the mirror image of invariant representation in sensory cortex. On the sensory side, a wide variety of input patterns can activate a stable cell assembly that represents some abstract pattern (your friend's face, your sunglasses). On the motor side, a stable cell assembly representing some abstract motor command (catching a ball, signing your name) is able to express itself using a wide variety of muscle groups and respecting a wide variety of other constraints. This symmetry between perception and action is what we should expect if, as Mountcastle proposed, the cortex runs a single basic algorithm in all areas.(HAWKINS, BLAKESLEE, 2004)

Sua proposta, portanto, seria replicar este comportamento de predição em software. As redes neurais biológicas espontaneamente detectam padrões e se auto-

³⁵ *invariant representations*

organizam em representações invariantes que dão origem à percepção, ao comportamento, aos pensamentos, à imaginação e à consciência. Sendo assim, ao imitar a natureza, seria possível trazer todos estes resultados para redes neurais em silício. O desafio estaria em criar padrões que representem os objetos e que possam ser manipulados sem perder a identidade. Seu objetivo não é reproduzir o conjunto completo da inteligência humana, mas o mecanismo essencial que possa ser adaptado a inúmeras tarefas específicas.

Avançando na perspectiva de modelo de Inteligência baseado em predição, Marcus Hutter propôs, em 2000, um modelo formal de agente³⁶ inteligente que, argumenta, é o sistema mais inteligente possível, independente do ambiente ou objetivo. Este modelo foi denominado AIXI. Para ele, a característica que une todas as perspectivas, ou ao menos a maior parte, sobre a inteligência em geral é a capacidade de maximizar uma função de utilidade para um objetivo. No mesmo veio que Hawkins, equaciona inteligência e predição, desenvolvendo uma teoria que aproxima Teoria de Decisão e Teoria da Informação. Sua intenção é desenvolver uma AI universal capaz de aprender através da observação do ambiente e, eventualmente, abstrair um modelo capaz de prever o próximo estado do sistema(HUTTER, 2000, 2003).

A teoria de decisão estuda como agentes realizam suas decisões em uma situação na qual os recursos são escassos. Dados estes recursos, que podem ser o tempo ou a quantidade de informações disponíveis sobre o mundo, por exemplo, o agente procura tomar sempre a melhor decisão. Para tanto, o agente precisa formular um modelo do mundo, isso é dizer, como o agente acredita que o mundo se comporta.

A teoria da indução de Solomonoff é a base sobre a qual Hutter vai modelar a inteligência. Dadas as limitadas observações sobre o mundo às quais o agente é exposto, um grande número de hipóteses sobre como o mundo funciona competem, nenhuma parecendo melhor que a outra. Hutter pede que todas estas hipóteses sejam consideradas. Isto remete ao Princípio das Múltiplas Explicações,

³⁶ Estamos utilizando “agente” como um ente abstrato, pode ser uma pessoa, uma empresa, um animal. Com isto pretende-se extrair todas as particularidades e ficar apenas com o essencial para o argumento.

originalmente atribuído a Epicuro. Legg enuncia o princípio como “mantenha todas as explicações consistentes com os dados” (LEGG, 2008. Tradução nossa.). Em um passo seguinte apela para a navalha de Occam para elencar por critério de probabilidade quais as hipóteses mais consistentes, “dentre todas as hipóteses consistentes com as observações, a mais simples é a mais provável” (LEGG, 2008. Tradução nossa.).

Um último elemento é necessário para se completar o modelo de Hutter, que é a inferência indutiva, a qual é dependente da crença do agente sobre o estado e funcionamento do mundo antes da observação. Para isto, o autor incorpora a Regra de Bayes ao modelo, organizando todas as hipóteses em uma distribuição de probabilidade. A cada nova evidência, a distribuição é alterada para refletir este novo dado. A probabilidade atrelada a cada hipótese antes de qualquer observação é chamada *probabilidade a priori*³⁷, e pode ser entendida como o conjunto de *vieses* construído no agente. Neste sentido, podemos pensar que a navalha de Occam é um viés do modelo, atribuindo maior probabilidade às explicações mais simples. Por exemplo, dada a sequência 2, 4, 6, 8, qual o próximo número?

Hipótese A: $f(x)=2x, x \in \mathbb{N}$

Hipótese B: $f(x)=2x^4-20x^3+70x^2-98x+48, x \in \mathbb{N}$

Se utilizarmos a navalha de Occam, diremos que a hipótese A é a mais provável, e apostaremos que o número 10 é o próximo. Caso o resultado seja 58, será necessário atualizar o modelo após a observação.

Para se pensar este modelo, usando um exemplo adaptado de Legg, podemos montar um cenário no qual o agente precisa adivinhar o número de caras em um lançamento de duas moedas. Os possíveis resultados seriam, nenhuma cara, uma cara ou duas caras:

Resultados possíveis: {0,1,2}

O sistema precisa ter uma forma de recompensa para que haja interesse no agente de acertar o resultado no jogo, digamos, ficar com as moedas. A estratégia, que reflete o conhecimento do mundo do agente, instrui este a sempre apostar em

³⁷ Frequentemente abreviado por *prior*.

uma cara a cada lançamento se as moedas forem honestas. Mas, não o sendo, apostar em 0 ou 2. Logo temos,

o: observação {sucesso, insucesso}

r: resultado {cara, coroa}

a: ação {apostar 0, apostar 1, apostar 2}

Fazemos do agente uma função π , que toma o histórico como argumento e devolve a próxima ação: $\pi(a_3|o_1r_1a_1o_2r_2)$. O ambiente, denotado como μ , é a distribuição de probabilidade de o_kr_k dado o histórico de interação $(o_{k-1}r_{k-1}a_{k-1})$. Um agente irá utilizar o histórico de moedas lançadas para avaliar se as moedas são honestas ou tendem mais a um resultado que outro, e escolher a ação ótima em função disto, de forma a melhor reproduzir μ (LEGG, HUTTER, 2007b).

O mais simples contexto para se situar este modelo é a predição binária, pois estabelece uma limitada gama de ações, predizer “1” ou “0”, em um universo de qualquer complexidade possível. Funcionaria como um microcosmos para se testar a eficiência deste agente. Este problema de predição foi explorado por Solomonoff ainda em 1964, em seu influente artigo *A formal theory of inductive inference*. O que se procura é um programa que compute o próximo dígito dada uma sequência. Digamos, uma sequência de zeros extremamente longa, haverá um infinito número de programas capazes de reproduzir este padrão. Porém, pela navalha de Occam, atribuímos uma probabilidade maior ao mais simples³⁸. Isto é, a complexidade dos dados se reflete na complexidade da hipótese que o agente utiliza para realizar a predição.

Mas é possível, é claro, que um conjunto de dados não possua informação o suficiente para se extrair um padrão. O conjunto pode ser incompleto a tal ponto que a inferência é impossível. Há também a possibilidade de que um conjunto seja fácil de ser gerado, mas a decodificação seja extremamente taxativa tornando o padrão, na prática, impossível de ser reconhecido. Desta forma, um agente ótimo π é aquele capaz de gerar dados que reproduzem o ambiente μ de maneira mais fiel possível.

³⁸ Isto é, baixa complexidade de Kolmogorov.

Uma limitação desta abordagem formal é o fato de entrar no domínio do não computável e, portanto, resta a dúvida se uma versão prática deste modelo de inteligência é alcançável.

4.5 *Conclusão*

De acordo com a modelo proposta de Hutter e Legg, o agente inteligente é aquele capaz de prever os eventos no mundo, mesmo que seu modelo do mundo seja imperfeito. Em um ambiente suficientemente complexo, o agente não tem como inferir um modelo totalmente preciso – sobretudo se fizer, ele mesmo, parte deste ambiente. Isto levaria a um evidente paradoxo: posto que constantemente interferem nos próximos eventos, portanto, precisam também prever a si mesmos. Sendo assim, um mecanismo de tomada de decisões em meio a informações incompletas e imprecisas se faz necessário para os seres vivos prosperarem. A inteligência nos organismos biológicos é, portanto, uma importante ferramenta de sobrevivência.

Seguindo esta linha de pensamento, não é de se surpreender, que há uma pressão seletiva forte para se adquirir mais desta característica. Seria de se esperar que tal capacidade fosse amplamente distribuída, de maneira espectral, nos seres vivos, ou, pelo menos, no reino animal. No entanto, quando nos reportamos à inteligência humana, usamos definições e jargões que não incluem, ou são pouco apropriadas, para descrever a inteligência dos outros animais, que muitas vezes sequer admitimos como inteligência, usando termos menos assertivos como comportamento, cognição. Conjecturamos, que isto acontece não por animais não possuírem características que possamos chamar de inteligentes, mas porque existe um uso muito específico do termo quando se refere ao humano, o qual acreditamos ser demasiado estrito para ser generalizado. Este aspecto reflete um excepcionalismo dos atributos humanos tão marcante que podemos chamar de viés antropocêntrico.

Prosseguindo o raciocínio, notamos que testes de IQ medem apenas parte do espectro do que o termo “inteligência” nos remete. Não há dúvida de que a inteligência seja útil no espaço social e cultural humanos, mas tudo indica que estamos acostumados a exagerar nas implicações ontológicas e axiológicas disso, ou

seja, este atributo é tão útil para os humanos como qualquer outra inteligência biológica parece ser: um conjunto de ferramentas para lidar com aspectos comuns do ambiente. Aspectos estes aos quais os organismos foram submetidos ao longo de milhões de anos de evolução, produzindo meios para a sobrevivência e sucesso reprodutivo.

Poderíamos incorrer em alguma forma de simplismo se não notássemos que organismos biológicos possuem um conjunto de técnicas para prosperar nos diversos ambientes nos quais se encontram que são em, em conjunto, muito mais poderosos que a inteligência medida pelos testes. Com base em tudo que foi exposto, consideramos mais apropriado chamar todos estes comportamentos que mostram flexibilidade na resolução de problemas, adaptabilidade ao ambiente, criatividade, abstração para resolver problemas, ou atingir um objetivo, de *inteligentes*. E este agente, seja uma humano, um animal, planta ou uma máquina, chamaremos de inteligente ou dotado de inteligência. Neste mesmo movimento, absorvendo o aspecto de natural ou artificial como um acidente, meramente dependente do contexto sobre o qual nos debruçamos e, portanto, não essencial para se entender, definir ou desenvolver inteligência.

A matemática, especificamente a capacidade de modelar matematicamente fenômenos, é fundamental para aprimorar a compreensão do mundo. E, neste sentido, a Ciência da Computação nos fornece uma ferramenta muito precisa para tal: a capacidade de simular os fenômenos e, aqui, em específico, de simular a inteligência. Idealmente gostaríamos de, a partir dos princípios mais básicos, derivar conceitos mais complexos e, em algum ramo deste desenvolvimento, chegar à inteligência humana. Esta abordagem permitiria mapear o espaço multidimensional da inteligência onde poderíamos localizar todas as possíveis manifestações dela. Isto, se a hipótese de que uma inteligência geral seja possível, ao menos abstratamente no infinito. Outros ramos levariam a inteligências diferentes, mais apropriadas a outros ambientes, reais ou hipotéticos. Conjuntos avançados desses ramos poderiam ser chamados de superinteligências. Nessa lógica, o que separa os elementos pertencentes a estes conjuntos daqueles que não pertencem é a capacidade de comportamento inteligente. Porém, uma inteligência geral, embora possa ser imaginada abstratamente, pode não ser realizável, nem em silício, nem nos

organismos biológicos, pois sempre é possível que um conjunto de dados não seja compatível com as abordagens conhecidas.

Isto é, se os recursos forem escassos no ambiente e a informação incompleta, o agente inteligente é aquele capaz de produzir um comportamento tal, dentro de tempo relevante, de sua capacidade e com seus instrumentos, que é melhor que uma ação meramente aleatória. Desta forma, a inteligência, em geral, pode ser entendida como um processo de otimização que maximiza a eficiência de um indivíduo, conjunto de indivíduos ou sistema, em um dado ambiente.

4.6 *Discussão*

Há duas perspectivas mutuamente excludentes para se desenvolver AI. Uma delas tem na natureza e na neurociência uma forma de inspiração, embora seus modelos, na prática, guardem pouca similaridade com os mecanismos biológicos – que permanecem, contudo, pouco compreendidos. O contraexemplo que seus detratores comumente evocam é o de se inventar o avião se inspirando nos pássaros. Em sua defesa, muitas abordagens nos vários campos da IA, assim como áreas contíguas, nasceram da imitação do que se observa já no mundo: otimizações baseadas em colônia de insetos, redes neurais, computação evolutiva, dentre os mais óbvios. Mas também, após um certo método heurístico se estabelecer, este toma um rumo próprio, com pés muito mais firmes na Matemática e na Ciência da Computação que na observação da natureza que lhe deu origem.

A outra perspectiva, defendida por Rich Sutton na sua postagem de *blog* em março de 2019, intitulado *The Bitter Lesson*³⁹, é a de que os algoritmos são secundários: o que importa é a disponibilidade de recursos computacionais. A razão pela qual certos problemas não puderam ser resolvidos, seja no final dos anos cinquenta, seja hoje, não é decorrente da falta de um algoritmo inteligente. Talvez os algoritmos que temos em cada época sejam tão inteligentes quanto eles podem ser nas condições limitadas de processamento computacional no qual foram desenvolvidos e implementados. Ou, por um outro ângulo, é necessário maior poder computacional para se desenvolver melhores algoritmos.

³⁹ <http://www.incompleteideas.net/IncIdeas/BitterLesson.html>

Este posicionamento conta com um embasamento é sólido, os algoritmos disponíveis no início das pesquisas em IA eram altamente especializados, minuciosamente ajustados, e precisavam ser executados por máquinas com recursos de memória e velocidade de processamento extremamente limitadas. Conforme o desenvolvimento trouxe processadores mais rápidos, seguindo a Lei de Moore, a generalidade dos algoritmos também aumentou, e cada vez mais menos intervenção humana é necessária para se atingir maior eficiência. Ao contrário, adicionar o conhecimento e experiência humana aos algoritmos é nocivo ao desenvolvimento da área, privilegiando apenas ganhos a curto prazo: a *força bruta* a longo prazo é muito mais poderosa.

De fato, os avanços recentes refletem exatamente isto. Os mais avançados algoritmos possuem mínima intervenção humana, é melhor ensinar um computador a jogar Go ou Xadrez contra si mesmo do que carregar o mesmo com imensas bibliotecas de aberturas e jogos notáveis. O que não deve surpreender, pois em poucos dias um computador consegue experienciar mais partidas que a humanidade já jogou em toda sua história. É mais eficiente dar todo conteúdo da internet para um computador devorar e descobrir por si mesmo, do que detalhadamente codificar as regras da gramática. Em verdade, humanos já nascem com aptidão para usar a linguagem, no entanto, é possível que não entendamos como ela funciona o suficiente para podermos programá-la em um computador. A visão computacional é um dos mais interessantes exemplos disso, pensemos em como o reconhecimento facial era ineficiente quando dependia de estabelecer relações entre características da face como a distância entre os olhos, ângulo com o nariz e a boca. Esta abordagem apenas retardou o desenvolvimento, desviando investimentos para uma técnica que, em breve, seria descartada sem nunca ter atingido a maturidade.

O progresso da AI se mostra principalmente nos algoritmos que usam busca e aprendizagem, pois estes são os que mais aproveitam os recursos computacionais que vêm se tornando disponíveis. Em certo sentido, seria como dizer que inteligência é busca e aprendizagem.

Marvin Minsky, ao contrário, acreditava que a inteligência não é uma única capacidade, mas um conjunto de adaptações biológicas que se acumularam durante a evolução e, portanto, não faz sentido procurar um único algoritmo ou método. Isto

é, quanto mais maneiras de se abordar um problema, mais inteligência o indivíduo manifesta. Assim, a AI que se entrincheira em uma única abordagem representa apenas parte do que a inteligência pode ser.

Arriscamos a seguir uma leitura da trajetória do desenvolvimento da AI para a partir dela estimar, em linhas gerais, o futuro panorama de pesquisa (*Figura 10*). As primeiras explorações, comumente denominada AI Simbolista⁴⁰, manifesta menor flexibilidade, o que era inerente às técnicas utilizada: os pesquisadores de então, nos anos 50 e 60 do século XX, procuravam codificar o conhecimento e técnicas conhecidas para solução de problemas através de algoritmos especializados. Não era comum se pensar em agentes. O sistema tomava o problema, executava o algoritmo, e devolvia a resposta. Um exemplo é o SAINT, *Symbolic Automatic Integrator*, de James Slagel, que resolvia problemas de Cálculo. Este tipo de técnica, frequentemente, não é mais entendido como parte da AI, entretanto, é muito empregada em máquinas e robôs simples, além de sistemas especialistas.

Em um nível maior de complexidade estariam os métodos que se utilizam de meta-heurística. Entretanto, estas técnicas são raramente incluídas na AI atualmente, sendo mais comumente associadas às técnicas de busca e otimização, embora tenha suas origens nas técnicas simbolistas. A diferença mais marcante em relação à AI simbolista, é o uso explícito de agentes para se resolverem os problemas, o que confere mais flexibilidade. Um exemplo é a ACO, *Ant Colony Optimization*, que cria agentes que simulam o comportamento de formigas em busca de alimento (APÊNDICE C). Agentes interagem entre si e podem se adaptar dinamicamente às situações que se transformam no decorrer do tempo.

⁴⁰ Ou como denominado por John Haugeland, *Good Old-fashioned Artificial Intelligence*.

Em terceiro nível, situamos as redes neurais. Estas são capazes de formular um modelo do mundo a partir dos dados que são utilizados no treinamento. Podem exercer um comportamento adversarial para aumentar a eficiência da fase de aprendizagem. As redes neurais solucionam os problemas através da extração de informação significativa de exemplos, como padrões que se repetem parcialmente. São utilizadas para processamento de linguagem, visão computacional e robótica avançada.

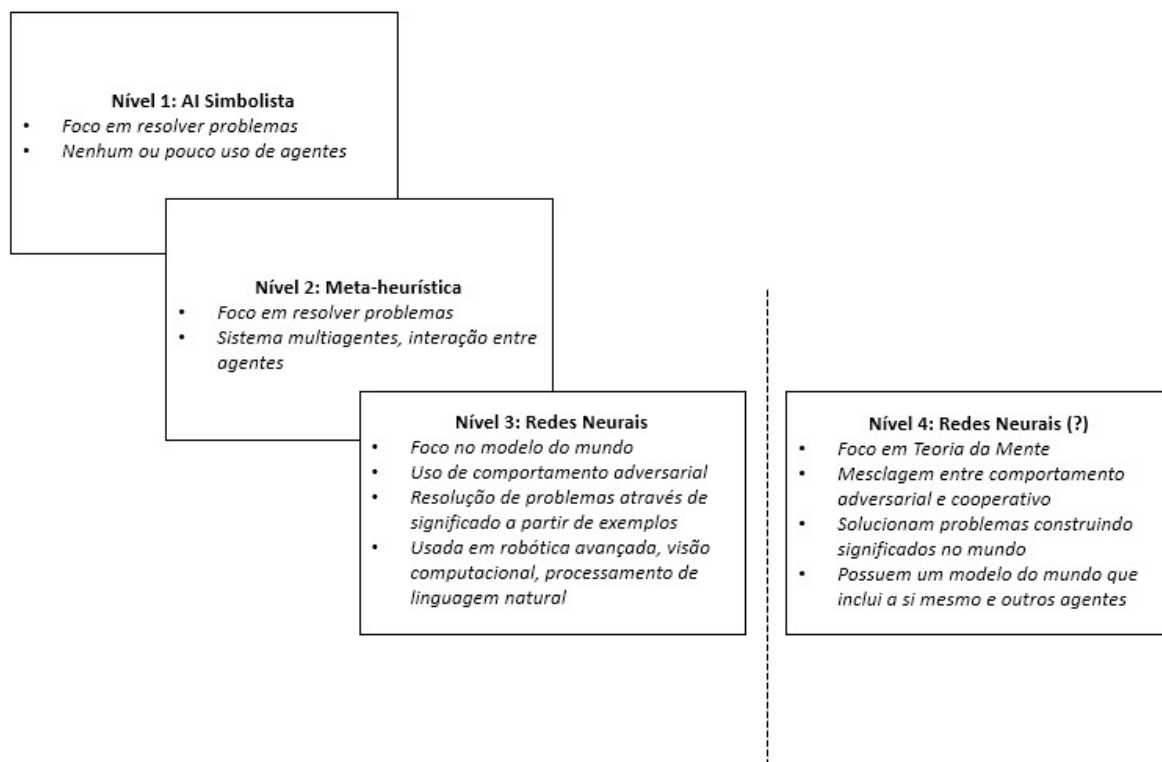


Figura 90: Proposta de perfil para pesquisa em inteligência de máquinas

O nível mais alto nesta escala então seria ocupado por técnicas que focassem em agentes com teoria da mente. Em um nível mais simples pode-se imaginar agentes que são capazes tanto de cooperar quanto competir, levando em consideração a ação e intenções dos demais agentes. Provavelmente isto precisa ser construído sobre redes neurais incorporando Teoria dos Jogos. Estes agentes precisam atribuir significado às coisas do mundo, não apenas padrões simples. Isto será importante para balancear a procura pelo menor resultado a curto prazo versus um ganho maior a longo prazo. Para tanto, o agente não apenas teria um modelo do mundo, mas teria a si mesmo inserido neste mundo, com seus interesses e valores.

Esta abordagem em níveis tem um resultado bastante interessante ao nosso ver, pois o agente começa a se tornar cada vez mais similar a um organismo biológico, apesar das restrições do ambiente simulado. Entendemos que, desta maneira, há certa convergência que pode não ser acidental, mas necessária: um agente mais complexo pode ser aplicado a problemas mais complexos, enquanto um agente simples não seria capaz de decodificar esse ambiente. Ademais, isto pode apontar para a emergência da consciência na evolução natural. Não, obviamente, na gama de sentidos que a linguagem adquiriu para se referir à consciência humana, que é por demais impreciso e antropocêntrico, mas em uma forma de “proto-consciência” mais afim às bactérias, talvez. O que queremos sublinhar aqui é que da mesma maneira como vemos a inteligência como herança evolutiva, não há razão para imaginar que com a consciência seria diferente, possuindo, portanto, vários graus e nuances típicas de cada espécie. O termo, que é vago e impreciso, poderia então ser descrito de forma funcional, ou seja, de que maneira *alguma* consciência seria melhor que *nenhuma* para conferir vantagem evolutiva? O que, temos a impressão, pode estar relacionado a uma melhor capacidade de organização dos objetivos, o que resultaria em um uso mais eficiente dos recursos disponíveis. Em outras palavras, a consciência pode ser modelada como uma meta-heurística.

5 CONCLUSÃO

A história da invenção do computador tem várias facetas, a do mito de dar a vida para o inanimado, a do autoconhecimento, a da superação da condição humana, e a da dessacralização da mente, psique ou alma. Nenhuma dessas idealizações foi alcançado em sua totalidade, e continuam a nos motivar e nos assombrar. A computação tem aberto novos caminhos onde a matemática do contínuo está deixando de ser produtiva: a estatística, aos poucos, vem ganhando o espaço antes reservado às soluções analíticas na ciência, a simulação acaba sendo mais poderosa que as equações, e os modelos – *theoryless* – oriundos do Machine Learning podem ter mais poder preditivo que as teorias estabelecidas. Isto, levado a um extremo, pode apontar que a cognição humana não é, talvez, o melhor instrumento para entender o mundo. Contra intuitivamente, o mundo mesmo talvez seja o melhor instrumento para entender a si mesmo. Desenvolver, continuamente, métodos cada vez mais eficientes na extração de padrões, a partir das regularidades do mundo, parece ser a maneira mais eficiente de alavancar o conhecimento científico.

Embora uma definição de inteligência seja uma tarefa demasiado complexa e muitas vezes resistente a delimitações, nossos esforços valeram a pena no sentido de termos feito emergir, neste processo, algumas características que podemos entender como um mínimo comum do conceito, como a capacidade de adaptação ao ambiente e a capacidade de prever como o mundo irá se comportar em um momento futuro. Além disso, dentro de uma perspectiva evolutiva, a inteligência está diretamente ligada aos desafios que organismos encontram e precisam superar para sobrevivência. Em um mundo em que cada ação tem um custo e um risco associado, a inteligência se apresenta como a capacidade de maximizar as chances de o organismo prosperar. Neste sentido, a inteligência representa ou é representada por um algoritmo – ou ecorritmo - de otimização, no âmbito de cada problema enfrentado.

REFERÊNCIAS

- ABBAGNANO, N. **Dicionário de Filosofia**. Tradução: Alfredo Bosi. 4a. ed. São Paulo, Martins Fonte, 2003.
- ADLEMAN, L. M. "Molecular Computation of Solutions to Combinatorial Problems", **Science**, v. 266, n. 11 November, p. 1021–1024, 1994. .
- BABBAGE, C., "Of the Analytical Engine". **Passages from the Life of a Philosopher**, 1st. ed. London, Longman, 1864. p. 112–141. Disponível em: <http://www.gutenberg.org/files/57532/57532-h/57532-h.htm>.
- BARRICELLI, N. A. "Numerical testing of evolution theories part I: Theoretical introduction and basic tests", **Acta Biotheoretica**, v. 16, n. 1, p. 69–98, 1962. .
- BASTIAN, B., LOUGHNAN, S., HASLAM, N., *et al.* "Don't mind meat? the denial of mind to animals used for human consumption", **Personality and Social Psychology Bulletin**, v. 38, n. 2, p. 247–256, 2012. DOI: 10.1177/0146167211424291. .
- BONNER, A. **The art and logic of Ramon Llull: a user's guide**. 1st. ed. Leiden, Brill, 2007.
- BONNER, A., "What Was Llull Up To?". In: FIDORA, A., SIERRA, C. (Org.), **Ramon Llull: From the Ars Magna to Artificial Intelligence**, 1st. ed. Barcelona, Consejo Superior de Investigaciones Científicas, 2011. p. 5–24. Disponível em: <http://www.iiia.csic.es/library>.
- BOOLE, G. **An investigation of the laws of thought on which are founded the mathematical theories of logic and probabilities / by George Boole**. 1st. Digit ed. London, Walton and Maberly, 1854.
- BORING, E. G., "Intelligence as the Tests Test It.". **Studies in individual differences: The search for intelligence.**, East Norwalk, Appleton-Century-Crofts, 1961. p. 210–214. DOI: 10.1037/11491-017. Disponível em: <http://content.apa.org/books/11491-017>.
- BRATANOVA, B., LOUGHNAN, S., BASTIAN, B. "The effect of categorization as food on the perceived moral standing of animals", **Appetite**, v. 57, n. 1, p. 193–196, 2011. DOI: 10.1016/j.appet.2011.04.020. .
- BRÄUER, J., HANUS, D., PIKA, S., *et al.* "Old and new approaches to animal cognition: There is not "one cognition"", **Journal of Intelligence**, v. 8, n. 3, p. 1–25, 2020. DOI: 10.3390/jintelligence8030028. .
- BUCK, G. H., HUNKA, S. M. "W. Stanley Jevons, Allan Marquand, and the Origins of Digital Computing", **IEEE Annals of the History of Computing**, v. 21, n. 4, p. 21–25, 1999. DOI: 1058-6180/99. .
- BUCKNER, C. "Morgan's Canon, meet Hume's Dictum: Avoiding anthropofabulation in cross-species comparisons", **Biology and Philosophy**, v. 28, n. 5, p. 853–871, 2013. DOI: 10.1007/s10539-013-9376-0. .
- CHAITIN, G. **Proving Darwin: making biology mathematical**. New York,

Pantheon Books, 2012.

CONWAY-MORRIS, S. "The Cambrian "explosion" of metazoans and molecular biology: Would Darwin be satisfied?", **International Journal of Developmental Biology**, v. 47, n. 7–8, p. 505–515, 2003. DOI: 10.1387/19. .

COPELAND, B. J., "The Church-Turing Thesis". In: ZALTA, E. N. (Org.), **The Stanford Encyclopedia of Philosophy**, Spring 201 ed. [S.l.], Metaphysics Research Lab, Stanford University, 2019. .

COUTURAT, L. **The Logic of Leibniz**. Tradução: Donald Rutherford, R. Timothy Monroe. Online, UC San Diego, 1997. Disponível em: <http://philosophyfaculty.ucsd.edu/faculty/rutherford/Leibniz/contents.htm>.

CROSSLEY, J. N., "Ramon Llull's Contributions to Computer Science". In: FIDORA, A., SIERRA, C. (Org.), **Ramon Llull: From the Ars Magna to Artificial Intelligence**, 1st. ed. Barcelona, Consejo Superior de Investigaciones Científicas, 2011. p. 39–60. Disponível em: <http://www.iii.csic.es/library>.

DA SILVA, J. J. "Godel and transcendental phenomenology", **Revue Internationale de Philosophie**, v. 59, n. 234, p. 553–574, 2005. .

DAN BILEFSKY. "New Heyday for an Old Giant, the Golem". **The New York Times**, Prague, Europe, 10 maio 2009. Disponível em: https://www.nytimes.com/2009/05/11/world/europe/11golem.html?hp=&page_wanted=print. Acesso em: 31 dez. 2020.

DARWIN, C. **On the Origin of Species**. 1st. Revis ed. Oxford UK, Oxford University Press, 2009.

DAVIS, M. **Engines of logic: mathematicians and the Origin of th Computer**. 2nd. ed. New York, W. W. Norton & Company, 2001.

DAWKINS, R. **The Blind Watchmaker**. 1st. ed. New York, Norton, 1985.

DE WAAL, F. **The Brains of the Animal Kingdom**. 22 mar. 2013. Wall Street Journal. Disponível em: <https://www.wsj.com/articles/SB10001424127887323869604578370574285382756>. Acesso em: 2 jan. 2021.

DEARY, I. J., LAWN, M., BARTHOLOMEW, D. J. ""A conversation between Charles Spearman, Godfrey Thomson, and Edward L. Thorndike: The International Examinations Inquiry Meetings 1931-1938": Correction to Deary, Lawn, and Bartholomew (2008).", **History of Psychology**, v. 11, n. 3, p. 163–163, ago. 2008. DOI: 10.1037/1093-4510.11.3.163. .

DENNETT, D. C. **Darwin's Dangerous Idea: Evolution and the Meanings of Life**. [S.l: s.n.], 1995. Disponível em: <http://books.google.com/books?id=0vLaAAAAMAAJ>.

DESCARTES, R., "Meditations On First Philosophy". In: HALDANE, E., ROSS, G. R. T. (Org.), **Philosophical Works**, Dover Publ ed. New York, Cambridge University Press, 1955. p. 155.

DICKE, U., ROTH, G. "Neuronal factors determining high intelligence", **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 371,

n. 1685, 2016. DOI: 10.1098/rstb.2015.0180. .

DINHAM, S. "Primary Schooling In Australia: Pseudo-Science Plus Extras Times Growing Inequality Equals Decline", **Researchgate.Net**, p. 1–14, 2014. Disponível em:
https://www.researchgate.net/profile/Stephen_Dinham/publication/294874584_Primary_Schooling_In_Australia_Pseudo-Science_Plus_Extras_Times_Growing_Inequality_Equals_Decline/links/56c4e8ab08aea564e304d93e/Primary-Schooling-In-Australia-Pseudo-Science-Plus-E.

DOMINGOS, P. **The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World**. 1st. ed. New York, Basic Books, 2015.

DORIGO, M., MANIEZZO, V., COLORNI, A. "Ant System: Optimization by a Colony of Cooperating Agents", **IEEE Transactions on Systems, Man, And Cybernetics - Part B (Cybernetics)**, v. 26, n. 1, p. 1–13, 1996. DOI:
<https://doi.org/10.1109/3477.484436>. .

DORIGO, M., STÜTZLE, T. **Ant Colony Optimization**. 1st. ed. Cambridge, Massachusetts, MIT Press, 2004.

DYSON, G. **Turing's Cathedral: The Origins of the Digital Universe**. 1st. ed. New York, Vintage Books, 2012.

FINNEY, E. M., FINE, I., DOBKINS, K. R. "Visual stimuli activate auditory cortex in the deaf", **Nature Neuroscience**, v. 4, n. 12, p. 1171–1173, 2001. DOI:
 10.1038/nn763. .

GALILEI, G., "Excerpts from The Assayer". **Discoveries and Opinions of Galileo**, 1st. ed. New York, Random House LLC, 1957. p. 229–280.

GAULKROGER, S. **Descartes: an intellectual biography**. 1st. ed. New York, Oxford University Press, 1995.

GODFREY-SMITH, P. **Cephalopods and the evolution of the mind. Pacific Conservation Biology**. [S.l: s.n.]. , 2013

GOULD, S. J. **The Mismeasure of Man**. 1 Ed. ed. London, Penguin Books, 1992.

GRANDIN, T. **Thinking in Pictures: My life with autism**. 2nd Ed. ed. New York, Random House LLC, 2006.

GREGORY, R. ., ZANGWILL, O. L. (Org.). **The Oxford Companion to the Mind**. 1st. ed. Oxford, Oxford University Press, 1987.

HAWKINS, J., BLAKESLEE, S. "On intelligence", **Neural Networks**, p. 272, 2004. Disponível em: www.onintelligence.com.

HERCULANO-HOUZEL, S. "The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost", **Proceedings of the National Academy of Sciences**, v. 109, n. Supplement_1, p. 10661–10668, 2012. DOI:
 10.1073/pnas.1201895109. .

HERCULANO-HOUZEL, Suzana, MANGER, P. R., KAAS, J. H. "Brain scaling in mammalian evolution as a consequence of concerted and mosaic changes in numbers of neurons and average neuronal cell size", **Frontiers in Neuroanatomy**, v. 8, n. August, p. 1–28, 2014. DOI: 10.3389/fnana.2014.00077. .

- HESPOS, S. J., VAN MARLE, K. "Physics for infants: Characterizing the origins of knowledge about objects, substances, and number", **Wiley Interdisciplinary Reviews: Cognitive Science**, v. 3, n. 1, p. 19–27, 2012. DOI: 10.1002/wcs.157. .
- HOBBS, T., "Computation or Logic, Chapter I.". In: MOLESWORTH, W., THULCYDIDES, H. (Org.), **The English works of Thomas Hobbes of Masmesbury, vol. I**, London, J. Bohn, 1839. p. 1–12.
- HOPKINS, W. D., RUSSELL, J., MCINTYRE, J., *et al.* "Are chimpanzees really so poor at understanding imperative pointing? Some new data and an alternative view of canine and ape social cognition", **PLoS ONE**, v. 8, n. 11, p. 1–6, 2013. DOI: 10.1371/journal.pone.0079338. .
- HORN, J. L. "Organization of Abilities and the Development of Intelligence", **Psychological Review**, v. 75, n. 3, p. 242–259, 1968. DOI: 10.1037/h0025662. .
- HUSSERL, E. **Ideas: general introduction to pure phenomenology**. 1st. ed. Abington, Routledge, 2013.
- HUTTER, M. "A gentle introduction to the universal algorithmic agent AIXI", **Artificial General Intelligence**, n. January, 2003. Disponível em: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.11.8797%5Cnhttp://www.researchgate.net/publication/228851452_A_gentle_introduction_to_the_universal_algorithmic_agent_AIXI/file/9fcfd513fb2bb4148d.pdf.
- HUTTER, M. "A Theory of Universal Artificial Intelligence based on Algorithmic Complexity", 2000. Disponível em: <http://arxiv.org/abs/cs/0004001>.
- JENSEN, A. R. "Galton's legacy to research on intelligence", **Journal of Biosocial Science**, v. 34, n. 2, p. 145–172, 2002. DOI: 10.1017/s0021932002001451. .
- JOHNSON, R. C., MCCLEARN, G. E., YUEN, S., *et al.* "Galton's Data a Century Later", **American Psychologist**, v. 40, n. 8, p. 875–892, 1985. DOI: 10.1037/0003-066X.40.8.875. .
- JOHNSTONE, M. A., "On 'Logos' in Heraclitus". **Oxford Studies in Ancient Philosophy, Volume 47**, [S.l: s.n.], 2016. p. 1–30. DOI: 10.1093/acprof:oso/9780198722717.003.0001.
- KANT, I. **Critique of Pure Reason**. Tradução: Paul Guyer, Allen W. Wood. 1st. ed. Cambridge, Cambridge University Press, 1998.
- KARI, L., DALEY, M., GLOOR, G., *et al.* "How to compute with DNA". 1738, 1999. **Anais [...]** Berlin, Heidelberg, Springer Berlin Heidelberg, 1999. p. 269–282. DOI: 10.1007/3-540-46691-6_21.
- KAUFMAN, A. S., RAIFORD, S. E., COALSON, D. L. **Intelligent testing with the WISC-V**. Hoboken, NJ, Wiley, 2016.
- KIM, E., TOOLE, B. A., ERIC KIM, E., *et al.* "Ada and the First Computer", **Scientific American**, v. 280, n. May, p. 76–81, maio 1999. DOI: 10.1038/scientificamerican0599-76. Disponível em: <http://www.nature.com/doi/10.1038/scientificamerican0599-76>.
- KLEMENT, K. C., "New Logic and the Seeds of Analytic Philosophy: Boole, Frege". In: SHAND, J. (Org.), **A Companion to Nineteenth-Century Philosophy**, 1st. ed.

Hoboken, NJ, Wiley-Blackwell, 2019. p. 454–479. DOI: 10.1002/9781119210054.ch17.

KOLMOGOROV, A. N. "Three approaches to the quantitative definition of information", **International Journal of Computer Mathematics**, v. 2, n. 1–4, p. 157–168, 1968. DOI: 10.1080/00207166808803030. .

LAZZARONI, M., MARSHALL-PESCINI, S., MANZENREITER, H., *et al.* "Why do dogs look back at the human in an impossible task? Looking back behaviour may be over-interpreted", **Animal Cognition**, v. 23, n. 3, p. 427–441, 2020. DOI: 10.1007/s10071-020-01345-8. Disponível em: <https://doi.org/10.1007/s10071-020-01345-8>.

LEGG, S. **Machine Super Intelligence**. 2008. 184 f. University of Lugano, 2008. Disponível em: <http://www.umiacs.umd.edu/~jbg/docs/nips2009-rtl.pdf%5Cnhttp://www.aclweb.org/anthology/W12-3304%5Cnhttp://portal.acm.org/citation.cfm?doid=1964858.1964870%5Cnhttp://dl.acm.org/citation.cfm?id=1964870%5Cnhttp://arxiv.org/abs/1409.2329%5Cnhttp://www.arxiv>.

LEGG, S., HUTTER, M. "A Collection of Definitions of Intelligence", **arXiv preprint**, p. 1–12, 2007a. Disponível em: <http://arxiv.org/abs/0706.3639>.

LEGG, S., HUTTER, M. "Universal intelligence: A definition of machine intelligence", **Minds and Machines**, v. 17, n. 4, p. 391–444, 2007b. DOI: 10.1007/s11023-007-9079-x. .

LEIBNIZ, G. W. **Die philosophischen Schriften von Gottfried Wilhelm Leibniz, vol. 7**. Facsimile ed. Hildesheim, Verlag, 1978.

LENSKI, R. E., OFRIA, C., PENNOCK, R. T., *et al.* "The evolutionary origin of complex features", **Nature**, v. 423, n. May, p. 139–144, 2003. Disponível em: <papers2://publication/uuid/CBB2DEEB-8D92-4816-94B3-8459AB8408DC>.

LOUGHNAN, S., HASLAM, N., BASTIAN, B. "The role of meat consumption in the denial of moral status and mind to meat animals", **Appetite**, v. 55, n. 1, p. 156–159, 2010. DOI: 10.1016/j.appet.2010.05.043. .

MACLEAN, E. L., HERRMANN, E., SUCHINDRAN, S., *et al.* "Individual differences in cooperative communicative skills are more similar between dogs and humans than chimpanzees", **Animal Behaviour**, v. 126, p. 41–51, 2017. DOI: 10.1016/j.anbehav.2017.01.005. Disponível em: <http://dx.doi.org/10.1016/j.anbehav.2017.01.005>.

MAYR, E. **The Growth of Biological Thought: diversity, Evolution and Inheritance**. 1st. ed. Cambridge, MA, Harvard University Press, 1982.

MCCARTHY, J., MINSKY, M. L., ROCHESTER, N., *et al.* "A proposal for the Dartmouth summer research project on artificial intelligence", **AI Magazine**, v. 27, n. 4, p. 12–14, 2006. .

MCCORDUCK, P. **Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence**. 2nd Editio ed. Natick, MA, A K Peters, Ltd., 2004.

MENABREA, L. F., LOVELACE, A. A. **Sketch of The Analytical Engine invented by**

- Charle Babbage.** 1842. Bibliothèque Universelle de Genève. Disponível em: <https://www.fourmilab.ch/babbage/sketch.html>.
- METTRIE, J. O. de La. **Man a Machine.** Chicago, The Open Court Publishing Co., 1912. Disponível em: <http://www.archive.org/details/manmachine00lame>.
- MIKLÓSI, Á., PONGRÁCZ, P., LAKATOS, G., *et al.* "A comparative study of the use of visual communicative signals in interactions between dogs (*Canis familiaris*) and humans and cats (*Felis catus*) and humans", **Journal of Comparative Psychology**, v. 119, n. 2, p. 179–186, 2005. DOI: 10.1037/0735-7036.119.2.179. .
- MILLÁN, J. B.-S. "Lingua characterica and calculus ratiocinator: the Leibnizian background of the Frege-Schröder polemic", **The Review of Symbolic Logic**, p. 1–36, 29 jun. 2020. DOI: 10.1017/S175502031900025X. Disponível em: https://www.cambridge.org/core/product/identifier/S175502031900025X/type/journal_article.
- MINSKY, M. L. **The Society of Mind.** 1ed. Paper ed. New York, Simon & Schuster, 2007.
- MONOD, J. **Chance and Necessity: An Essay on the Natural Philosophy of Modern Biology.** 1st. ed. New York, Knopf, 1971.
- MORAVEC, H. **Mind Children: The Future of Robot and Human Intelligence.** 1 ed. ed. Cambridge, Massachusetts, Harvard University Press, 1988. Disponível em: <http://www.archive.org/details/mindchildrenfutu00mora>.
- MOUNTCASTLE, V. B., EDELMAN, G. M., "An organizing principle for the cerebral function: the unit module and the distributed system". **The Mindful Brain: Cortical Organization and the Group-selective Theory of Higher Brain Function**, 1st. ed. Cambridge, Massachusetts, The MIT Press, 1978. p. 7–50.
- MULAİK, S. A. **Foundations of Factor Analysis.** 2nd. ed. Boca Raton, FL, Chapman and Hall/CRC, 2010.
- NEUMANN, J. von. "First draft report on EDVAC", **IEEE Ann. Hist. Comput.**, v. 15, n. 4, p. 27–75, 1993. .
- NEUMANN, J. von, KURZWEIL, R., CHURCHLAND, P., *et al.* **The Computer and the Brain.** 3rd. ed. New Haven, Yale University Press, 2012.
- NEWELL, A., SIMON, H. A. "The logic theory machine--A complex information processing system", **IRE Transactions on information theory**, v. 2, n. 3, p. 61–79, 1956. .
- NILSSON, N. J. **The Quest for Artificial Intelligence.** 1st. ed. Cambridge, Cambridge University Press, 2010.
- O'BRIEN, M. D. "Nerve endings: the discovery of the synapse.", **Journal of the Royal Society of Medicine**, v. 99, n. 6, p. 322, jun. 2006. .
- OLAZARAN, M. "A Sociological Study of the Official History of the Perceptrons Controversy", **Social Studies of Science**, v. 26, n. 3, p. 611–659, 29 ago. 1996. DOI: 10.1177/030631296026003005. Disponível em: <http://journals.sagepub.com/doi/10.1177/030631296026003005>.
- OPEN SCIENCE COLLABORATION. "The Reproducibility Project", **Implementing**

Reproducible Research, p. 1–83, 2014. Disponível em:
<http://dx.doi.org/10.1201/b16868-14>.

PAYNE, H. L., LYNCH, G. F., ARONOV, D. "Precise spatial representations in the hippocampus of a food-caching bird", **bioRxiv**, 2020. DOI:
 10.1101/2020.11.27.399444..

PÉREZ FRAGA, P., GERENCSÉR, L., LOVAS, M., *et al.* "Who turns to the human? Companion pigs' and dogs' behaviour in the unsolvable task paradigm", **Animal Cognition**, n. 0123456789, 2020. DOI: 10.1007/s10071-020-01410-2. Disponível em: <https://doi.org/10.1007/s10071-020-01410-2>.

PETZOLD, C. **The Annotated Turing: A Guided Tour Through Alan Turing's Historic Paper on Computability and the Turing Machine**. Indianapolis, John Wiley & Sons, 2008.

PICCININI, G. "The first computational theory of mind and brain: A close look at McCulloch and Pitts's "logical calculus of ideas immanent in nervous activity"", **Synthese**, v. 141, n. 2, p. 175–215, 2004. DOI:
 10.1023/B:SYNT.0000043018.52445.3e. .

PINKER, S. "The Blank Slate, the Noble Savage, and the Ghost in the Machine", 1999. .

RHODIUS, A. **Argonautica**. 1st new ed ed. Cambridge, MA, Harvard University Press, 2009. Disponível em:
https://www.loebclassics.com/view/apollonius_rhodes-argonautica/2009/pb_LCL001.5.xml.

ROE, A. W., PALLAS, S. L., KWON, Y. H., *et al.* "Visual projections routed to the auditory pathway in ferrets: Receptive fields of visual neurons in primary auditory cortex", **Journal of Neuroscience**, v. 12, n. 9, p. 3651–3664, 1992. DOI:
 10.1523/jneurosci.12-09-03651.1992. .

ROSENBLATT, F. "The Design of an Intelligent Automaton", **U.S. Office of Naval Research**, v. 6, n. 2, p. 7, 1958. .

RUSSELL, S. J., NORVIG, P. **Artificial Intelligence: a modern approach**. 4 Ed. ed. Hoboken, NJ, Pearson, 2020.

SALES, T., "Lull as a Computer Scientist, or Why Lull was one of us". In: FIDORA, A., SIERRA, C. (Org.), **Ramon Lull: from the Ars Magna to Artificial Intelligence**, 1st. ed. Barcelona, Consejo Superior de Investigaciones Científicas, 2011. p. 25–38. Disponível em: <http://www.iiia.csic.es/library>.

SCHWARTZ, O. **In the 17th Century, Leibniz Dreamed of a Machine That Could Calculate Ideas**. 4 nov. 2019. IEEE Spectrum Tech Talk. Disponível em:
<https://spectrum.ieee.org/tech-talk/artificial-intelligence/machine-learning/in-the-17th-century-leibniz-dreamed-of-a-machine-that-could-calculate-ideas>. Acesso em: 23 jan. 2021.

SEARLE, J. R. "Is the brain's mind a computer program?", **Scientific American**, v. 262, n. Jan, p. 26–37, 1990. Disponível em:
http://sils.shoin.ac.jp/~gunji/AI/CR/Is_The_Brains_Mind_A_Computer_Program.pdf. Acesso em: 6 ago. 2014.

SEARLE, J. R. **The Philosophy of Mind (Lectures)**. . Springfield, VA, Teacghing Co. , 1996

SHANNON, C. E. "A Mathematical Theory of Communication", **Bell System Technical Journal**, v. 27, n. 3, p. 379–423, jul. 1948. DOI: 10.1002/j.1538-7305.1948.tb01338.x. Disponível em:
<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6773024>.

SMULLYAN, R. M. **Godel's Incompleteness Theorems**. 1st. ed. New York, Oxford University Press, 1992.

SNIR, S., YOHAY, B. "Prokaryotic evolutionary mechanisms accelerate learning", **Discrete Applied Mathematics**, v. 258, p. 222–234, 2019. DOI: 10.1016/j.dam.2018.10.044. Disponível em:
<https://doi.org/10.1016/j.dam.2018.10.044>.

STRONGMAN, L. **The anthropomorphic bias: How human thinking is prone to be self-referential**. . [S.l: s.n.], 2007. Disponível em:
<http://repository.openpolytechnic.ac.nz/handle/123456789/1245>.

TERMAN, L. M., MERRILL, M. A. **Stanford-Binet intelligence scale: Manual for the third revision, form IM**. Boston, Houghton Mifflin, 1960.

TERMAN, L. M., ODEN, M. H. **The gifted group at mid-life: Thirty-five years' follow-up of the superior child. Volume V**. California, Stanford University Press, 1959. Disponível em: <https://archive.org/details/giftedgroupatmid011505mbp>.

THE NEW YORK TIMES. "New Navy Device Learns by Doing; Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser". **The New York Times**, New York, NY, 8 jul. 1958. , p. 25–25 Disponível em:
<http://www.nytimes.com/1958/07/08/archives/new-navy-device-learns-by-doing-psychologist-shows-embryo-of.html>. Acesso em: 28 dez. 2020.

TOOLE, B. A. **Ada, The Enchantress of Numbers**. 1. ed. Sausalito, CA, Critical Connection, 1992.

TRUITT, E. R. **Medieval Robots: mechanism, magic, nature, and art**. Philadelphia, University of Pennsylvania Press, Inc., 2015.

TURING, A. M. "Computing Machinery and Intelligence", **Mind**, v. LIX, n. 236, p. 433–460, 1 out. 1950. DOI: 10.1093/mind/LIX.236.433. Disponível em:
<https://academic.oup.com/mind/article/LIX/236/433/986238>.

TURING, A. M. "On computable numbers, with an application to the Entscheidungsproblem", **Proceedings of the London Mathematical Society**, v. 42, n. 2, p. 230–265, 1936. Disponível em:
<http://plms.oxfordjournals.org/cgi/reprint/s2-42/1/230>.

VALIANT, L. **Probably Approximately Correct: Nature's Algorithms for Learning and Prospering in a Complex World**. 1st. ed. New York, Basic Books, 2013.

VAN HEIJENOORT, J. "Logic as calculus and logic as language", **Synthese**, v. 17, n. 1, p. 324–330, 1967. DOI: 10.1007/BF00485036. .

WANG, P. "On Defining Artificial Intelligence", **Journal of Artificial General**

Intelligence, v. 10, n. 2, p. 1–37, 2019. DOI: 10.2478/jagi-2019-0002. .

WARNE, R. T. "An Evaluation (and Vindication?) of Lewis Terman: What the Father of Gifted Education Can Teach the 21st Century", **Gifted Child Quarterly**, v. 63, n. 1, p. 3–21, 2019. DOI: 10.1177/0016986218799433. .

WATERHOUSE, L. **Multiple intelligences, the Mozart effect, and emotional intelligence: A critical review. Educational Psychologist**. [S.l.], Lawrence Erlbaum Associates, Inc. . Disponível em: https://www.tandfonline.com/doi/abs/10.1207/s15326985ep4104_1. Acesso em: 28 jan. 2021. , set. 2006

WATSON, P. **Ideas: A History of Thought and Invention**. [S.l.: s.n.], 2006.

WIENER, N. **Cybernetics: or the Control and Communication in the Animal and the Machine**. 2nd. ed. Cambridge, Massachusetts, MIT Press, 1961.

WOOD, G. **Café Curio: Gaby Wood - Inventing Robots - What's On**. . Online, Camden Arts Centre. Disponível em: <http://camdenartscentre.org/whats-on/view/eve-bl-10/>. , 2012

WOOD, G. **Edison's Eve: a magical history of the quest for mechanical life**. 1st. ed. New York, A.A. Knopf, 2002.

APÊNDICE A

P vs. NP e o Problema do Caixeiro Viajante¹

Há problemas para os quais encontrar uma solução ótima ou exata requer muitos passos, que podem ser entendidos como intensos computacionalmente ou intensos no uso de memória, mas ainda assim são essencialmente tratáveis. Um exemplo seria dizer se certo número é primo (AGRAWAL, KAYAL, *et al.*, 2004). Trata-se de um problema de decisão, isto é, a resposta é um “sim” ou “não”. Para este caso específico existem algoritmos para os quais é garantido obter tal resposta em tempo polinomial, o que significa dizer que se precisarmos decidir sobre um número primo muito grande, basicamente bastaria um computador rápido para se obter o resultado. No entanto, estima-se que a maioria dos possíveis problemas não possui solução algorítmica, como o *Halting Problem*, explorado por Alan Turing em seu artigo seminal (TURING, 1936). Entre estes dois extremos há diversas classes de problemas que, ao mesmo tempo que não são passíveis de serem resolvidos em um tempo razoável, ainda possuem, em tese, solução algorítmica (PETZOLD, 2008). Isto é, uma Máquina de Turing seria perfeitamente capaz de encontrar a solução, dado tempo e papel infinitos para realizar as operações necessárias (GÖDEL, CHURCH, *et al.*, 1965).

Abordando como classe, externamente teríamos um plano infinito O onde estariam todos os problemas ou funções possíveis. Um problema em O pode ser resolvido por alguma Máquina Oráculo em n passos (*Fig. 1*). Neste espaço encontraríamos uma classe de problemas R , os quais podem ser resolvidos por algoritmos recursivos em uma máquina de Turing, em tempo e memória infinitos (SOARE, 1987). Internamente temos os problemas na forma c^n , onde c é uma constante e n um polinômio, isto significa que o tempo de execução varia exponencialmente em função da entrada. Internamente a próxima classe é NP , estes podem, no entanto, ser resolvidos rapidamente por uma máquina não-determinística, mas uma máquina de Turing, necessita explorar todo o espaço de soluções sequencialmente, em grande parte dos casos demandando um tempo extraordinariamente longo (GÖDEL, CHURCH, *et al.*, 1965). A classe internamente seguinte, P , limita-se a problemas que podem ser resolvidos em tempo polinomial, n^c , por uma máquina de Turing. Esta é a classe dos problemas que são considerados tratáveis, no sentido de existir um algoritmo rápido para se chegar a solução (AARONSON, 2013, GÖDEL, CHURCH, *et al.*, 1965). Porém apenas um pequeno conjunto, P' , contém os problemas que, de fato, são viáveis de serem resolvidos de um ponto de vista prático.

Embora não se saiba se $P=NP$, é em geral presumido que este não é o caso. Um problema é dito NP-hard se for ao menos tão complexo quanto o mais complexo na classe NP. Isto implica, portanto, que um problema em EXP-hard é NP-hard, mesmo que não esteja

¹ Este texto é um excerto adaptado de KOVALSKI, E. L. **Algoritmos de Otimização na Alocação de Recursos em Datacenters**. 25 f. (Artigo). Programa de Pós-Graduação em Tecnologia da Informação, Universidade do Sul de Santa Catarina (UNISUL), 2018. Disponível em <https://riuni.unisul.br/handle/12345/5340?locale-attribute=en>

em NP. Cabe notar que, igualmente, não podemos afirmar $NP=EXP$ como verdadeiro ou falso (GÖDEL, CHURCH, *et al.*, 1965).

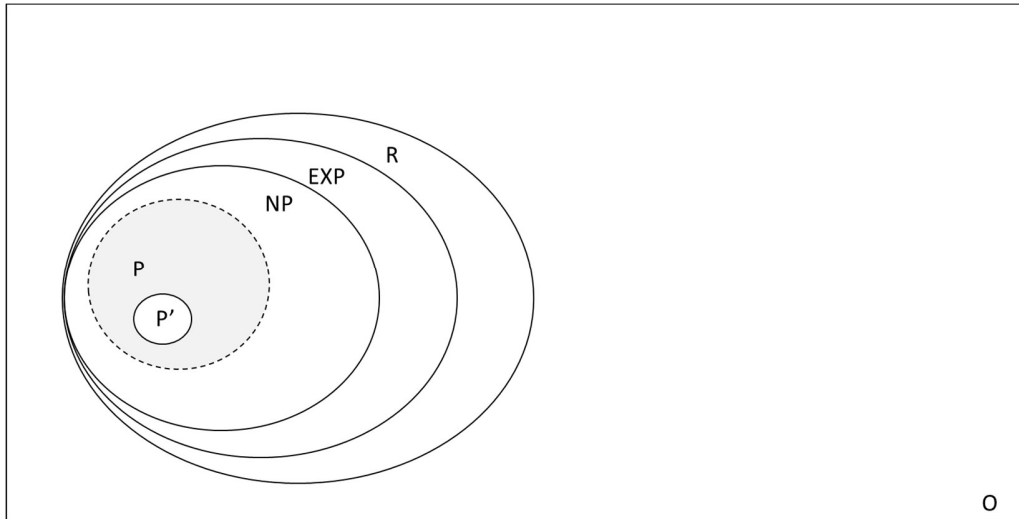


Figura 1: Classes de Problemas

Para se afirmar que um problema novo B é NP-hard, prova-se reduzindo um problema conhecido A NP-hard para ele, com isto significando que B é ao menos tão complexo quanto A. Se um problema A que é sabidamente hard na classe NP, pode ser reduzido para outro, problema B, da mesma classe, então B é dito NP-completo desde que esta redução se dê em tempo polinomial. Desta maneira, um problema NP-completo pode ser transformado em qualquer outro NP-completo e possuirá a mesma solução. Cabe lembrar que, de posse da solução a um problema NP-Hard, e que esteja ele mesmo em NP, verificar a correção do mesmo é possível em tempo polinomial (esta é, de fato, a definição de NP) (GÖDEL, CHURCH, *et al.*, 1965).

Problemas de decisão, como comentado acima, possuem uma resposta do tipo “sim” ou “não”. Neste sentido, não há muito que se possa fazer quanto a melhorar a solução de um problema de decisão NP-hard, ou o resultado é obtido e se pode certificar sua validade, ou não é obtido e se continua a explorar o espaço de pesquisa. Por outro lado, problemas de otimização procuram o melhor resultado para um dado problema, isto é, produz-se resultados parciais, que podem estar longe de serem ótimos, e tempo adicional de processamento é utilizado com a finalidade de melhorá-los. Isto abre a possibilidade de se deixar de lado a procura pelo resultado ótimo, já que este se mostra inviável, e desenvolverem-se técnicas para procurar um resultado que seja *bom o suficiente*. Em outras palavras, quando métodos exatos se mostram infrutíferos pode se lançar mão de métodos heurísticos para se chegar a uma resposta aproximada.

O Problema do Caixeiro Viajante, ou *Traveling Salesman Problem* (TSP) é apresentado como: dadas n cidades e suas distâncias entre si, encontrar o menor circuito que visita todas as cidades exatamente uma vez. Embora esteja relacionado a como encontrar o

Embora à primeira vista as questões de complexidade algorítmica façam parte de uma área da matemática que se apresenta como um campo abstrato o qual parece desvinculado dos problemas reais do cotidiano, a aplicabilidade de soluções para estas questões no ambiente de datacenter indicia a intensa conexão entre os avanços na pesquisa teórica e as melhorias nas aplicações práticas.

Referências:

AARONSON, S., "Why Philosophers Should Care About Computational Complexity". In: COPELAND, J. B. (Org.), **Gödel, Turing, Church, and beyond**, Cambridge, MA, MIT Press, 2013. p. 261. DOI: 10.1080/10463280903275375.

AGRAWAL, B. M., KAYAL, N., SAXENA, N. "PRIMES is in P (Original)", **Annals of Mathematics**, v. 160, n. 2, p. 781–793, 2004. Disponível em: <http://www.jstor.org/stable/3597229>.

FORTNOW, L., HOMER, S. "A Short History of Computational Complexity", **Bulletin of the EATCS**, v. 80, p. 95--133, 2003. Disponível em: [https://www.dropbox.com/sh/v4blv7sito3opzn/AAD3Ub6b5WlrZLIHSwyA191oa/A Short History of Computational Complexity - Fortnow%2C Homer - 2002.pdf?dl=0](https://www.dropbox.com/sh/v4blv7sito3opzn/AAD3Ub6b5WlrZLIHSwyA191oa/A%20Short%20History%20of%20Computational%20Complexity%20-%20Fortnow%20-%20Homer%20-%202002.pdf?dl=0).

GÖDEL, K., CHURCH, A., TURING, A. M., *et al.* **The undecidable: Basic papers on undecidable propositions, unsolvable problems and computable functions**. New York, Raven Press, 1965.

KARP, R. M., "Reducibility among combinatorial problems". **50 Years of Integer Programming 1958-2008: From the Early Years to the State-of-the-Art**, [S.l.: s.n.], 2010. p. 219–241. DOI: 10.1007/978-3-540-68279-0_8.

LIPTON, R. J. M., RICHARD, J., KARP, R. M., *et al.* "Turing machines that take advice", **L'Enseignement Mathématique**, v. 28, p. 191–209, 1982. DOI: 10.1053/j.pcsu.2014.01.013. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/24727221>.

PETZOLD, C. **The Annotated Turing: A Guided Tour Through Alan Turing's Historic Paper on Computability and the Turing Machine**. Indianapolis, John Wiley & Sons, 2008.

SCHRIJVER, A. "On the History of Combinatorial Optimization", **Handbooks in Operations Research and Management Science: Discrete Optimization**, v. 12, n. Till 1960, p. 1–57, 2005. .

SOARE, R. I. **Recursively enumerable sets and degrees: A study of computable functions and computably generated sets**. New York, Springer-Verlag, 1987.

TURING, A. M. "On computable numbers, with an application to the Entscheidungsproblem", **Proceedings of the London Mathematical Society**, v. 42, n. 2, p. 230–265, 1936. Disponível em: <http://plms.oxfordjournals.org/cgi/reprint/s2-42/1/230>.

APÊNDICE B

Mentes Alienígenas: sobre as possibilidades de inteligência extraterrestre.

The universe is a pretty big place. If it's just us, seems like an awful waste of space.

Contact (1997)

Introdução

O filme *Contact* (1997) tem o mérito, entre muitos, de trazer o debate sobre a vida inteligente extraterrestre, dentro de um âmbito científico e racional, para o público leigo. Deixam o palco os homenzinhos verdes e seus discos voadores e entram sinais de rádio em um universo vasto demais para ser transposto mesmo por futurísticas espaçonaves. Mas em seu lugar emergem questões sobre a origem da vida em outras partes do universo e sua evolução, sobre os desafios de espécies inteligentes para evitar a destruição por sua própria tecnologia ou a extinção por meios naturais, sobre os limites do conhecimento e do domínio sobre a matéria e energia e, também, sobre a ética.

Um desperdício de espaço

Space is big. Really big. You just won't believe how vastly hugely mind-bogglingly big it is. I mean, you may think it's a long way down the road to the chemist, but that's just peanuts to space.

Douglas Adams

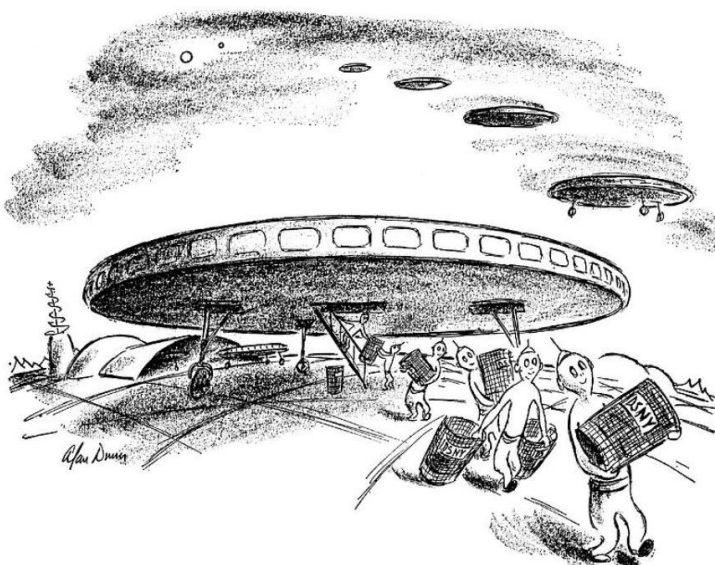


Figura 1: *New Yorker*, Maio de 1950. A sigla DSNY é para Department of Sanitation of NY

O físico Enrico Fermi, em 1950, em meio a um contexto de supostas observações objetos voadores não identificados (UFO) e a associação destas com o fenômeno do desaparecimento de lixeiras em Nova York (*Figura 1*) pelo cartunista

Alan Dunn, lança jocosamente a pergunta: onde está todo mundo? A abordagem desta questão, frequentemente retomada neste tipo de debate, ficou conhecida como o *Paradoxo de Fermi*, que pode ser enunciado como segue:

Há bilhões de estrelas similares ao nosso Sol na galáxia, e muitas destas são bilhões de anos mais antigas que nosso Sistema solar. Muito provavelmente algumas destas estrelas possuem planetas, e alguns deles são similares à Terra. Se o aparecimento de vida na Terra é típico, alguns destes planetas podem ter desenvolvido vida inteligente. Algumas destas civilizações podem ter desenvolvido capacidade de viagem interestelar, um passo que a humanidade está explorando no momento. Mesmo em velocidades mil vezes menores que a velocidade da luz, na escala de milhões de anos, a via láctea poderia ser atravessada de ponta a ponta. Então, onde está toda a vida inteligente extraterrestre?

O astrofísico Frank Drake (KURZWEIL, 2005), um dos fundadores do SETI (*Search for extraterrestrial intelligence*), tentou formular uma estimativa na forma de uma equação (*Quadro 1*), que poderia estabelecer uma faixa para a quantidade de espécies de vida inteligente que podemos esperar na galáxia.

Quadro 2: Equação de Drake

$$N = R^* \cdot f_p \cdot n_e \cdot f_l \cdot f_i \cdot f_c \cdot L$$

Onde:

N = Número de civilizações em nossa galáxia com as quais a comunicação seria possível em princípio

R^* = Média da taxa de formação de estrelas em nossa galáxia

f_p = Fração das estrelas que tem planetas

n_e = Média do número de planetas que podem suportar a vida por estrela

f_l = Fração dos planetas capazes de suportar a vida que de fato a desenvolvem

f_i = Fração dos planetas que desenvolvem vida inteligente

f_c = Fração de civilizações que desenvolvem tecnologia capaz de ser detectável através do espaço

L = Duração de tempo que estas civilizações permanecem emitindo sinais detectáveis através do espaço

Os resultados dependem principalmente do grau de otimismo utilizado, levando a resultados que variam de menor que 1 a 1.000.000 de civilizações ao nosso alcance. O melhor resultado, porém, é a estimativa de nosso grau incerteza.

Ray Kurzweill (2005) considera que o resultados que os programas do SETI encontraram são aproximadamente os esperados, isto é, provavelmente estamos sozinhos na galáxia. E seu argumento se baseia não apenas na não observação, mas na aceleração do desenvolvimento tecnológico (*Gráfico 1*) a partir do domínio do eletromagnetismo. Ao observar a revolução no uso de energia em pouco mais de 200 anos de história humana, percebe que este é mais aproximado pela modelagem através de uma curva exponencial, e isto permitiria fazer projeções que após

descoberto ondas de rádio, em cerca de 2000 anos tal civilização iria requerer mais energia anual que o total gerado pelo sol, e mais 1000 anos o total de energia da galáxia¹. Considerando a escala temporal que tratamos para o aparecimento da vida é da ordem de milhões de anos, alguma espécie com demandas energéticas desta natureza já teria sido observada neste ponto. Deveríamos neste caso esperar por uma exploração massiva de estrelas assim como de outras fontes de energia através da galáxia e também sinais de radiação eletromagnética produzida artificialmente seriam facilmente identificáveis, isso mesmo que equação de Drake nos dê $1 < N < 100$, no entanto o espaço parece totalmente intocado e silencioso.

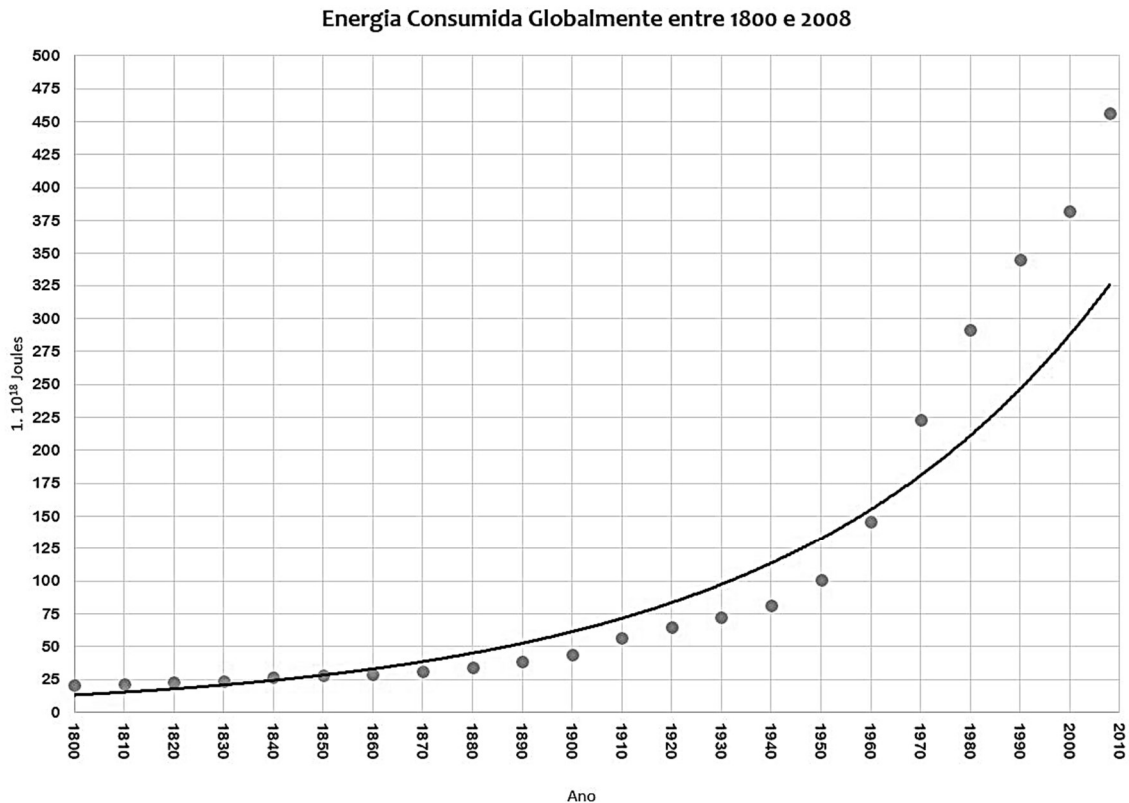


Gráfico 1: Consumo global de energia de 1800 a 2008. Dados em laranja, curva de regressão exponencial em azul. Fonte: (SMIL, 2010).

Onde está todo mundo?

Uma possível explicação para não nos depararmos com sinais de civilizações avançadas seria que estas passam por uma existência demasiado efêmera e nunca são capazes de desenvolverem-se o suficiente para deixarem o seu sistema solar. Esta abordagem pressupõe a existência de filtros em pontos da evolução, e a tecnologia pode ser um destes filtros levando a civilização a esgotar os próprios recursos energéticos antes de desenvolver meios de explorar fontes de energia alternativas. Isto colocaria um limite superior para a vida inteligente qual seria muito difícil escapar, similar a outros filtros que atuaram até o aparecimento de vida inteligente. O economista Robin D. Hanson (HANSON, 1998) criou uma lista que

¹ Estes dois pontos de requerimentos energéticos tomam como base a extrapolação do Gráfico 1, e grosseiramente tentam ser comparáveis com a Escala Kardashev para *Civilizações Tipo II* (4×10^{26} watts) e *Civilizações Tipo III* (4×10^{37} watts).

considera incompleta dos possíveis filtros que atuam limitando a expansão de vida inteligente, como nove passos evolutivos:

1. Estar em um Sistema Solar capaz de suportar a vida
2. Aparecimento de uma molécula reprodutiva capaz de preservar características hereditárias
3. Desenvolver vida unicelular simples
4. Desenvolver vida celular complexa, com organelas especializadas
5. Desenvolver reprodução sexual
6. Aparecimento de seres multicelulares
7. Aparecimento de seres capazes de usar ferramentas
8. Atingir o estado de desenvolvimento tecnológico similar ao que hoje possuímos
9. Colonização além do sistema solar de origem

Dentro de cada um destes passos, apenas uma ínfima parcela de ocorrências seria capaz de transpor para o passo seguinte, e talvez nenhuma ocorrência de vida tenha ainda superado o passo número oito em nossa galáxia. Talvez outras civilizações se tenham autodestruído quando descobriram a energia atômica, ou destruíram seu habitat por esgotamento de recursos, ou inventaram uma Inteligência Artificial que inadvertidamente eliminou seus inventores. Isto também significa que podemos estar nos deparando com este filtro neste momento evolutivo e há uma grande probabilidade de nossa tecnologia nos levar a extinção em poucos anos.

Porém, a não observação de empreendimentos de engenharia de escala galáctica² não serve como evidência de que tal filtro está em ação. Pode ser o caso de que somos de fato a primeira civilização a ser capaz de explorar a galáxia e, por sorte, todos os recursos estão disponíveis para nosso uso.

No entanto, há uma outra possibilidade, mais sóbria talvez, de que inteligências suficientemente mais desenvolvidas não têm interesse na espécie humana ou preferem não interferir no desenvolvimento das demais inteligências, agindo da mesma forma que um conservacionista ambiental, mas neste caso limitando a informação que flui em direção a espécies em desenvolvimento. Ou, em um sentido mais geral, civilizações avançadas simplesmente preferem não interagir com outras espécies e tomam medidas para evadirem-se à observação. Estas podem, afinal, representar uma possível ameaça existencial, assim espécies alienígenas evitam contato umas com as outras e deveríamos, portanto, ser mais cautelosos com nossas ondas de rádio.

Darwin, no espaço.

If I were to give an award for the single best idea anyone has ever had, I'd give it to Darwin.

Daniel Dennet

² Como as esferas ou conchas ao redor de estrelas que Freeman Dyson propôs como método para aproveitar o máximo da energia solar sem prejudicar a energia que naturalmente fluiria em direção ao planeta (TIMOFEEV, KARDASHEV, *et al.*, 2000).

Richard Dawkins nos remete a perceber como a teoria da evolução é brilhante pois explica como a complexidade é capaz de surgir do simples e Daniel Dennett expande sobre isso notando que nenhuma teoria rival é capaz de fazer o mesmo (DENNETT, 1995). Assim, é válido afirmar que provavelmente onde quer que a vida tenha surgido no universo, é através do processo evolutivo, isto é, seleção natural com acúmulo de mutações vantajosas, que a inteligência apareceu e foi moldada. Os limites do ambiente, confrontados às necessidades de obtenção de meios de sobrevivência assim como de ampliar o rol, de mutações vantajosas (por assim dizer, reprodução sexual), moldaram também os sentidos e, ultimamente, a cognição de toda espécie que atingiu uma civilização avançada através do universo. É possível, não obstante, que existam outros mecanismos, outras formas de “estar vivo” e de se adquirir níveis elevados de inteligência, mas estes seriam casos muito extremos para podermos levantar hipóteses.

Argumenta-se aqui que provavelmente após um certo nível de desenvolvimento tecnológico é alcançado, a espécie inteligente passa a entender o funcionamento de seu organismo e desenvolve próteses com a finalidade de reparar primeiro partes, órgãos e sentidos. Deste ponto, o próximo passo é utilizar a tecnologia para ampliar e aperfeiçoar a vida biológica, fazendo próteses mais eficientes que as partes naturais. Este processo de incorporação de tecnologia no biológico culminaria na superação completa do paradigma biológico e a adoção de outro processo evolutivo, a saber, através do Design Inteligente.

A Nave dos Argonautas

Este processo de Design pode não se limitar ao corpo físico, mas se estender à ampliação da capacidade cognitiva e mecanismos a mecanismos de preservação da consciência e memórias. Susan Schneider (SCHNEIDER, Susan, 2009) propõe o experimento mental do Paradoxo do Navio Teseu:

Imagine que o famoso navio que Teseu utilizou é guardado, como memento dos feitos do herói. Com o passar do tempo a madeira em certas partes começa a se decompor e é substituída, de tal forma que após cerca de um século todas as peças foram substituídas. Pergunta-se onde está o Navio de Teseu? Agora, usando a mesma lógica, imagine que com o tempo algumas funções mentais começam a mostrar sinais da idade, você vai ao médico e substitui partes do seu cérebro por equivalentes tecnológicos. Após o procedimento você se sente perfeitamente bem, e que nada de fato alterou em sua identidade. Você pergunta a seus amigos e familiares, e eles também não percebem nenhuma diferença. Após certo número de procedimentos, nenhuma parte de seu corpo permanece a originalmente biológica, você não sente qualquer diferença com respeito a sua identidade.

Este tipo de experimento mental pode vir a ser uma prática corriqueira em pouco tempo para a espécie humana. Avanços recentes na pesquisa biomédicas apontam para uma rápida incorporação de tecnologias tanto no sentido corretivo quanto ampliativo das funções cognitivas humanas (HAMPSON, SONG, *et al.*, 2018, MARKRAM, MULLER, *et al.*, 2015, SIMON, C LARSSON, *et al.*, 2015). Recentes avanços como estes representam significativo avanço no esclarecimento dos mecanismos responsáveis pela formação de memórias assim como nos sinais que o tecido

nervoso produz. É válido supor que o mesmo tenha se dado com outras espécies inteligentes.

Contatos Imediatos

Apesar de não ser esclarecido o substrato da consciência, é seguro afirmar ao menos dentro do paradigma científico, que esta decorre de processos ultimamente físicos (PENROSE, 1989). Independente da hipótese computacional ser correta, a passagem do biológico para um organismo pós-biológico pode incorporar uma ampla gama de sistemas que produzam as mesmas forças causais que dão origem ao pensamento consciente. Hoje os limites de nosso conhecimento sobre o funcionamento do cérebro, possibilidades tecnológicas e limitações computacionais não são suficientes para reproduzir uma mente consciente, ao que Nick Bostrom afirma que “persuasive arguments have been given to the effect that if technological progress continues unabated then these shortcomings will eventually be overcome” (BOSTROM, 2003), com o que concordamos. Uma civilização capaz de explorar além de seu sistema solar provavelmente já deve ter adquirido conhecimentos necessários para possibilitar a extensão da vida para além das limitações biológicas, haja vista os requisitos para longas viagens espaciais serem pouco compatíveis com a vida originada dentro da atmosfera de um planeta. Isto nos leva a primeira conclusão:

I - Os alienígenas que podemos vir a encontrar serão seres pós biológicos.

O recente avanço da computação a partir dos anos sessentas, e mais aceleradamente, a partir do uso de *deep learning*, proposto por Geoffrey Hinton apenas em 2006 (LECUN, BENGIO, *et al.*, 2015), e *generative adversarial networks*, por Ian Goodfellow em 2014 (SILVER, SCHRITTWIESER, *et al.*, 2017), nos apontam que independente de se reproduzir o pensamento consciente por meios tecnológicos (unicamente computacionais ou híbridos), a superinteligência é um fenômeno muito mais imediato³.

Superinteligência é uma forma de inteligência que supera a humana. Nick Bostrom distingue três tipos de superinteligência. Uma superinteligência pode ser superior à humana em termos de velocidade, isto é, embora não seja mais inteligente que um humano, devido à velocidade com que o algoritmo é executado pode realizar uma tarefa específica que um ser humano é capaz, porém o faz mais rápido. Superinteligência coletiva é quando o algoritmo em si não é muito inteligente, mas a organização de várias instâncias destes, operando em paralelo, superam a inteligência humana em âmbitos específicos do conhecimento. O último caso é o mais interessante, trata-se de uma superinteligência qualitativa, o que significa que não precisa ser mais rápida que a humana, mas é muito mais inteligente que o humano em todos os domínios (SCHNEIDER, Susan, 2016).

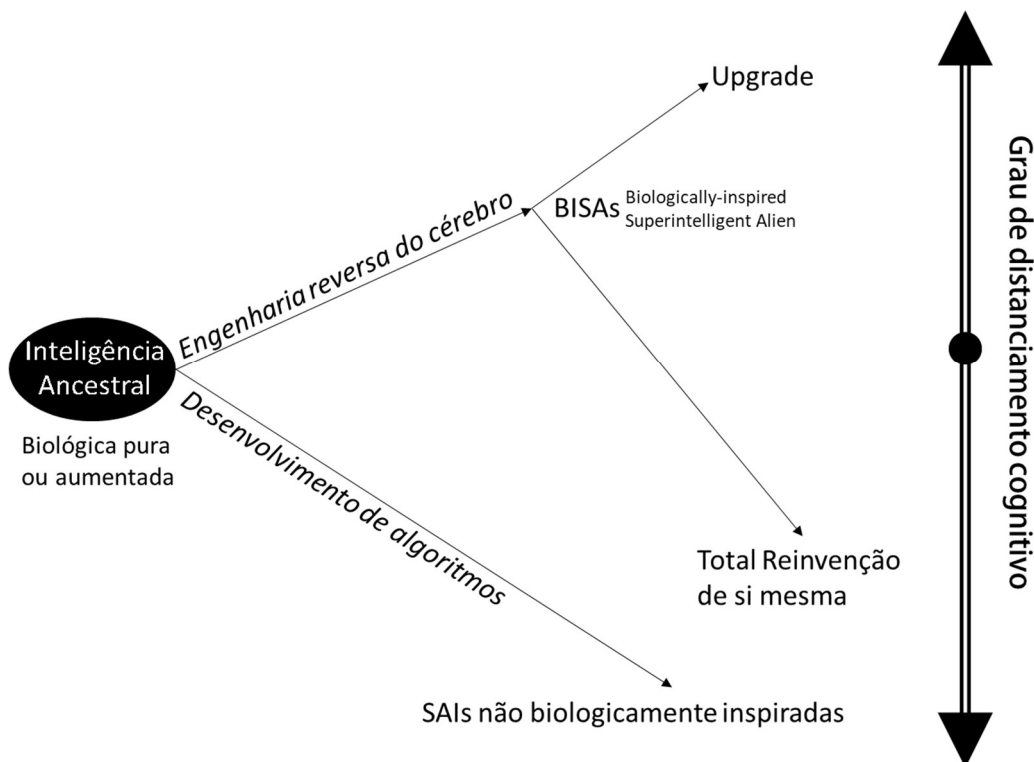
Uma forma de pensar a superinteligência qualitativa é o desenvolvimento de um programa de computador que é largamente é mais capaz que o humano de

³ De fato, em publicação recente na *Nature* com base na *engine* AlphaGo Zero há indícios de que apesar de estarmos apenas começando a desenvolver métodos de aprendizagem não supervisionada, os resultados iniciais já eclipsam todos os avanços obtidos com a aprendizagem supervisionada (SILVER, SCHRITTWIESER, *et al.*, 2017).

solucionar problemas, ao qual recorreríamos para descobrir a cura de uma doença, escrever o roteiro da próxima série de sucesso, governar uma nação, inventar a próxima geração de computadores superinteligentes. Outra abordagem é pensar na ampliação das capacidades cognitivas legadas pela evolução biológica. Neste caso poderíamos supor que ao substituir o hipocampo por uma prótese, esta permitiria preservar memórias indefinidamente e indexá-las de modo muito mais eficiente, ao substituir o processamento visual esta prótese nos daria a capacidade de imaginar espaços de n dimensões. No primeiro caso, a inteligência utilizada seria de uma estrutura muito diversa da biológica, não poderíamos entender os processos que permitem que um programa de computador chegue aos resultados, ainda que corretos. E cada Superinteligência Artificial (SAI) seria extremamente diversa de outra de acordo com a *tese de ortogonalidade* que discutiremos a seguir. Porém as superinteligências adquiridas por ampliação da capacidade cognitiva biológica seriam bastante similares entre si, seriam capazes de se comunicarem e trocarem conhecimentos. Defendemos que ambas as formas de superinteligência podem existir como inteligências alienígenas, porém apenas as baseadas em modelos biológicos seriam de interesse.

Little Green Man

Pela Tese de Ortogonalidade (BOSTROM, 2017), inteligência e objetivos são diversos, e uma superinteligência pode ter quase quaisquer objetivos independente do grau de superinteligência adquirido. Qualquer SAI provavelmente possui como objetivo a aquisição de recursos, aperfeiçoamento tecnológico, melhoria cognitiva, autopreservação e manutenção da integridade dos objetivos. Mas supomos que as *inteligências extraterrestres de inspiração biológica* (BISA) guardariam alguns dos objetivos específicas de seus ancestrais biológicos, como a procura por recursos, as



quais nos ajudariam a entender suas motivações e padrões de pensamento. Também seria mais fácil que BISAs pudessem ter mais disponibilidade para interagir com outras inteligências.

Figura 2: Possíveis Formas Avançadas de Inteligência Extraterrestre

Ademais, seria bastante possível que SAIs puras sejam não conscientes, seja pela incapacidade de reprodução da consciência no silício (ou seja qual for o substrato utilizado), seja porque a consciência seria uma característica supérflua. A consciência é um produto de evolução biológica que atende a necessidades específicas da vida. A consciência, por exemplo, permite focar a atenção para solucionar problemas que estão no limite de nossas capacidades cognitivas. Este não seria um problema para SAIs, devido ao enorme poder computacional adquirido e à possibilidade de gerar inúmeras instâncias de si mesma, virtuais ou físicas, para dar conta do que precisa computar. BISAs, por outro lado, é fácil supor que reteriam a capacidade da consciência. Isto nos leva à segunda conclusão:

II – As inteligências artificiais que podemos vir a encontrar são BISAs

Naturalmente, nem toda BISA irá preferir reter as mesmas características da espécie biológica ancestral, e algumas podem seguir a direção de reescreverem seu código de forma tão ousada que o vínculo se perderia. Porém, caminho da melhora através de correções pontuais e ampliação de capacidades manteria a compatibilidade entre BISAs de diferentes origens, o que pode ter algum valor.

Com base nestes pressupostos, podemos supor também que BISAs possuem uma forma de representação mental que é invariante com relação ao ponto de vista. Esta ferramenta é fundamental na cognição biológica, por razões óbvias (um predador de costas ou de lado ainda é o mesmo predador). Dentre as características a serem preservadas, esta tem um alto grau de possibilidade de estar entre elas. Outra, é a capacidade de representação combinatória e recursiva. E isto nos permitiria alguma forma de comunicação com estes seres (SCHNEIDER, Sunsan, 2009).

Conclusão

Time in the digital universe and time in our universe are governed by entirely different clocks. In our universe, time is a continuum. In a digital universe, time (T) is a countable number of discrete, sequential steps. A digital universe is bounded at the beginning, when T = 0, and at the end, if T comes to a stop. Even in a perfectly deterministic universe, there is no consistent method to predict the ending in advance. To an observer in our universe, the digital universe appears to be speeding up. To an observer in the digital universe, our universe appears to be slowing down.

George Dyson

Mas então, *onde está todo mundo?* É possível que uma rápida evolução possível para SAI leve rapidamente a um tipo de consciência qualitativamente tão diferente da consciência biológica que SAIs não reconheçam organismos biológicos

como conscientes. Talvez a escala de tempo com que operem tenha se tornado tão acelerada com relação ao tempo biológico que a vida pareça parada no tempo. Civilizações alienígenas avançadas podem ter deixado de explorar o universo e terem se movido, em parte ou totalmente, para universos mais perfeitos, universos simulados; pode ainda ser o caso de sermos a única espécie inteligente aqui em volta, no fim, por estarmos em um universo feito especificamente para simular a vida humana.

Referências:

- BOSTROM, N. "Are We Living in a Computer Simulation?", **The Philosophical Quarterly**, v. 53, n. 211, p. 243–255, 28 abr. 2003. DOI: 10.1111/1467-9213.00309. Disponível em: <http://pq.oxfordjournals.org/cgi/doi/10.1111/1467-9213.00309>.
- BOSTROM, N. **Superintelligence: Paths, dangers, strategies**. Oxford, Oxford University Press, 2017.
- DENNETT, D. C. **Darwin's Dangerous Idea: Evolution and the meaning of life**. 1. ed. London, Penguin Books, 1995.
- HAMPSON, R. E., SONG, D., ROBINSON, B. S., *et al.* "Developing a hippocampal neural prosthetic to facilitate human memory encoding and recall", **Journal of Neural Engineering**, v. 15, n. 3, 2018. DOI: 10.1088/1741-2552/aaaed7. .
- HANSON, R. "The great filter-are we almost past it", **preprint available at <http://hanson.gmu.edu/greatfilter.html>**, 1998. .
- KURZWEIL, R. **The Singularity is Near: When humans transcend biology**. New York, Penguin Books, 2005.
- LECUN, Y., BENGIO, Y., HINTON, G. "Deep learning", **Nature**, v. 521, n. 7553, 2015. DOI: 10.1038/nature14539. .
- MARKRAM, H., MULLER, E., RAMASWAMY, S., *et al.* "Reconstruction and Simulation of Neocortical Microcircuitry", **Cell**, v. 163, n. 2, 2015. DOI: 10.1016/j.cell.2015.09.029. .
- PENROSE, R. **The Emperor's New Mind: Concerning Computers, Minds and The Laws of Physics**. 1. ed. Oxford UK, Oxford University Press, 1989.
- SCHNEIDER, Susan. **Science fiction and Philosophy: from time travel to superintelligence**. [S.l.], Willey-Blackwell, 2009.
- SCHNEIDER, Susan, "Alien Minds". **Schneider, Susan. "Alien minds." Science Fiction and Philosophy: From Time Travel to Superintelligence**, 2. ed. [S.l.], Willey-Blackwell, 2016. p. 225–240.
- SILVER, D., SCHRITTWIESER, J., SIMONYAN, K., *et al.* "Mastering the game of Go without human knowledge", **Nature**, v. 550, n. 7676, 2017. DOI: 10.1038/nature24270. .
- SIMON, D., C LARSSON, K., NILSSON, D., *et al.* **An organic electronic biomimetic neuron enables auto-regulated neuromodulation**. [S.l.: s.n.], 2015. v. 71.
- SMIL, V. **Energy Transitions**. Santa Barbara, CA, Greenwood Publishing Group, 2010.
- TIMOFEEV, M. Y., KARDASHEV, N. S., PROMYSLOV, V. G. "A search of the IRAS database for evidence of Dyson Spheres", **Acta Astronautica**, v. 46, n. 10–12, p. 655–659, 2000. .

APÊNDICE C

Optimização Natural: o caso de formigas

Formigas ao explorarem o mundo em busca de alimento possuem um comportamento bastante eficiente para comunicar seus achados e recrutar as demais em seu formigueiro para explorarem os recursos. O objetivo é claro, conseguir a maior quantidade do alimento com o menor gasto energético possível, o que se traduz como um processo natural de otimização. A perspectiva que procuramos aqui clarificar é o aspecto de que simples regras de interação entre os indivíduos produzem uma forma de conhecimento que é útil para estes se desenvolverem.

As formigas utilizam-se de feromônios para a comunicação, além de movimentos corporais (Figura 1). Os detalhes exatos diferem de espécie para espécie, e outros insetos sociais possuem estratégias funcionalmente similares. No caso de abelhas, os significados parecem ser, ao menos em um caso, adquiridos socialmente (SU, CAI, *et al.*, 2008).

Workers of the ant *Camponotus socius* employ odor trails for either one of two purposes, to recruit nestmates to newly discovered food sources or to lead them to new nest sites (HÖLLDOBLER, 1971). Recruitment is specified by adding a wagging motion of the head while nest transfer is specified by a back-and-forth jerking motion of the entire body. (WILSON, 1975. p. 189)

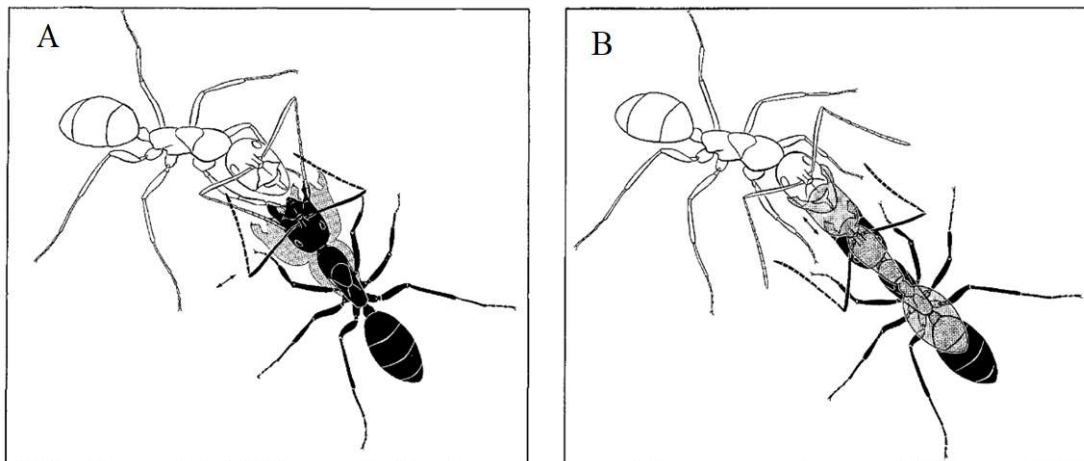


Figura 1: Sinalização composta em formigas operárias da espécie *Camponotus socius*. Além do feromônio deixado no ambiente, o movimento corporal adiciona informações ao sinal. Em **A** o movimento de recrutamento para busca de alimento e em **B** para indicar a transferência para outro local do formigueiro. Fonte: HÖLLDOBLER (1971).

A comunicação de agentes através de modificações no ambiente é denominada *estigmergia* quando esta proporciona a colaboração e a auto-organização. O termo foi

cunhado por Peirre-Paul Grassé (1895 – 1985) para se referir ao comportamento de cupins. Apesar de restrita comunicação direta e ausência de planejamento, insetos sociais produzem, através de estigmergia, estruturas complexas e engajam em comportamentos elaborados. Apesar de cada indivíduo possuir apenas um conhecimento limitado e local do estado do mundo.

Estes comportamentos podem ser simulados em software, adaptados e aprimorados para solucionar problemas humanos. As formigas, como indivíduos, possuem uma complexidade que está muito além de nosso entendimento atual e sua simulação realista não é possível dada a tecnologia existente no momento da escrita deste texto. Estamos procurando apenas os elementos mínimos essenciais para produção de um comportamento complexo. Assim, em nossa versão digital, a formiga pode ser descrita de forma simplificada como um agente em um espaço bidimensional com diversos obstáculos entre o formigueiro e o alimento. O obstáculo e o grau de dificuldade deste pode ser descrito como as diferentes distâncias a serem percorridas. O comportamento de busca de alimento será simplificado, usando apenas a comunicação intermediada pelo ambiente. Os agentes, neste esquema, a princípio exploram o espaço aleatoriamente, quando algum encontra o alimento este retorna marcando o caminho com o feromônio (Figura 2). O agente possui uma memória limitada, que é importante para evitar se perder em um *loop*. Após um certo tempo, o feromônio evapora. A presença e a quantidade de feromônio no trajeto moldam a probabilidade de o agente seguir certo caminho.

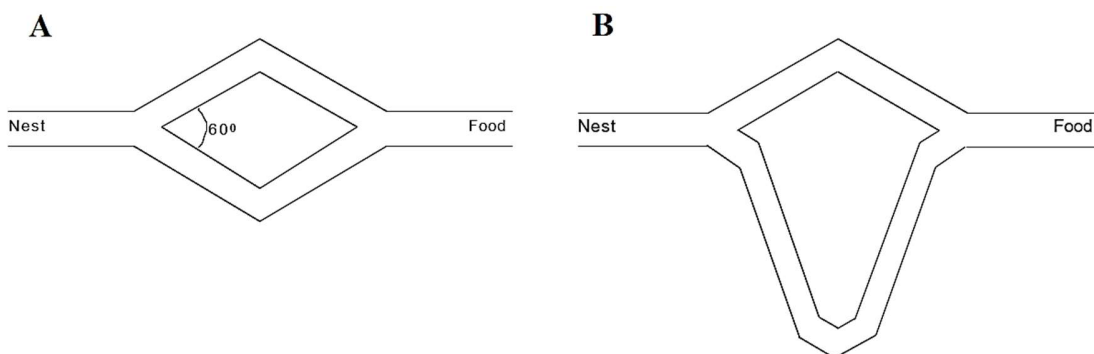


Figura 2: Na bifurcação em **A** a probabilidade do agente tomar cada caminho é a mesma. Em **B**, como um dos trajetos é mais longo, demora mais para acumular o feromônio, reduzindo gradualmente a probabilidade do próximo agente escolher este caminho. Fonte Dorigo e Stützle (2004)

Em formigas, não obstante, a evaporação não parece exercer um papel muito relevante. Após um caminho ter sido definido, elas apresentam resistência a usar um trajeto mais curto que só tenha se tornado disponível posteriormente. Porém o processo de evaporação, funcionando como um mecanismo de esquecimento, pode ser incorporado para regular a probabilidade de exploração de novas alternativas pelos agentes.

In fact, the great majority of ants choose the long branch because of its high pheromone concentration, and this autocatalytic behavior continues to reinforce the long branch, even if a shorter one appears. Pheromone evaporation, which could favor exploration of new paths, is too slow (...), which means that the pheromone evaporates too slowly to allow the ant colony to “forget” the suboptimal path to which they converged so that the new and shorter one can be discovered and “learned.”(DORIGO, STÜTZLE, 2004. p.5)

Um fenômeno que ocorre com formigas é, em algumas espécies, elas acabarem entrando em um ciclo quando sistematicamente reforçam com o feromônio um trajeto circular, ficando presas indefinidamente até exaurirem suas energias e perecerem (Figura 3). Isto é conhecido como espiral da morte (SCHNEIRLA, 1944). Cada agente, no entanto, possuindo uma certa quantidade de memória, proporcional à complexidade do espaço explorado, evita que a formação de um *loop* prejudique a convergência para o menor trajeto. A implementação deste mecanismo se dá através do registro de cada bifurcação pela qual passa. Ao encontrar o alimento, esse registro é processado removendo os *loops*, só então produzindo o caminho de volta no qual o feromônio será depositado.

Anteriormente mencionamos o uso de fitas de DNA como ferramenta computacional para se abordar o TSP. Este sistema de agentes inspirado no comportamento das formigas pode ser usado para o mesmo objetivo e, portanto, para qualquer problema que possa ser convertido para o TSP, incluindo problemas de agendamento, problemas de empacotamento, coloração de gráficos e aprendizado de máquinas, entre outros. Embora a eficiência deste tipo de abordagem em muitos casos seja insuficiente – mas nem sempre – frente a métodos desenhados especificamente para tais problemas¹, o que nos interessa é como regras simples dão origem a comportamentos e com grande capacidade de transformar o mundo.

¹ Em comparações com o Algoritmo de Dijkstra, A*, além de outras meta-heurísticas como Taboo Search e Simulated Annealing, dependendo muito do caso particular.

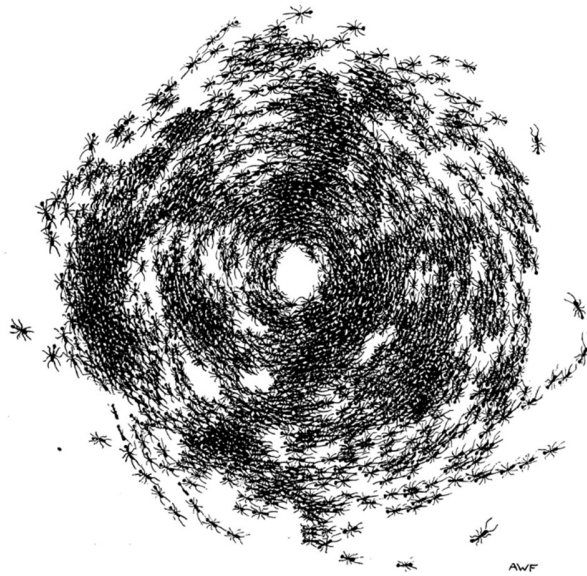


Figura 1: Desenho sobre uma foto de uma espiral da morte de formigas *Eciton praedator* com aproximadamente 15 cm de diâmetro. Fonte: SCHNEIRLA (1944)

Referências

- DORIGO, M., STÜTZLE, T. **Ant Colony Optimization**. 1st. ed. Cambridge, Massachusetts, MIT Press, 2004.
- HÖLLDOBLER, B. "Recruitment behavior in *Camponotus socius* (Hym. Formicidae)", **Zeitschrift für Vergleichende Physiologie**, v. 75, n. 2, p. 123-142, 1971. DOI: 10.1007/BF00335259. .
- SCHNEIRLA, T. C. "A unique case of circular milling in ants, considered in relation to trail following and the general problem of orientation", **American Museum Novitates**, n. 1253, p. 1-26, 1944. DOI: 10.1.1.174.1490. .
- SU, S., CAI, F., SI, A., *et al.* "East learns from west: Asiatic honeybees can understand dance language of European honeybees", **PLoS ONE**, v. 3, n. 6, 2008. DOI: 10.1371/journal.pone.0002365. .
- WILSON, E. O. **Sociobiology: the new synthesis**. 25th Anive ed. Cambridge, Massachusetts, Harvard University Press, 1975.