

MESTRADO EM FILOSOFIA
RAMO DE FILOSOFIA CONTEMPORÂNEA

O Robô Autodidata

Representação mental e autonomia cognitiva em inteligência artificial

Romeu Ivoleta Neto

M

2019



Romeu Ivoleta Neto

O Robô Autodidata

Representação mental e autonomia cognitiva em inteligência artificial

Dissertação realizada no âmbito do Mestrado em Filosofia – ramo de Filosofia

Contemporânea – orientada pelo Professor Doutor João Alberto Pinto

Faculdade de Letras da Universidade do Porto

outubro de 2019

O Robô Autodidata

Representação mental e autonomia cognitiva em inteligência artificial

Romeu Ivoleta Neto

Dissertação realizada no âmbito do Mestrado em Filosofia – ramo de Filosofia
Contemporânea – orientada pelo Professor Doutor João Alberto Pinto

Membros do Júri

Professor Doutor João Alberto Pinto
Faculdade de Letras – Universidade do Porto

Professor Doutor Mattia Riccardi
Faculdade de Letras – Universidade do Porto

Professora Doutora Sofia Gabriela Assis de Morais Miguens Travis
Faculdade de Letras – Universidade do Porto

Classificação obtida: 16 valores

Para Beatriz Ferri Conejo

Sumário

Declaração de honra	8
Agradecimentos	9
Resumo	10
Abstract.....	11
Lista de abreviaturas e siglas	13
Introdução	14
Observações sobre palavras-chave.....	16
Metodologia, objetivos e estrutura.....	17
Capítulo 1 – A representação mental e seus tipos	23
1.0 Representação mental «RM»	23
1.1 Tipo 1 – a representação mental simbólica (cognitivismo)	24
1.2 Tipo 2 – a representação mental subsimbólica (conexionismo)	29
1.3 Tipo 3 – a representação mental como ação incorporada (enativismo).....	39
Capítulo 2 – O tipo de RM necessária para a autonomia cognitiva em Hayy	44
2.1 A primeira fase cognitiva de Hayy	45
2.2 O experimento do bastão	46
2.3 O experimento do bastão sob as lentes do cognitivismo	48
2.4 O experimento do bastão sob as lentes do conexionismo.....	50
2.5 O experimento do bastão sob as lentes do enativismo.....	53
Capítulo 3 – As RM do tipo [Y] necessárias para a autonomia cognitiva em sistemas de IA	56
3.1 As criaturas de Rodney Brooks – inteligência sem representação	56
3.1.1 A RM do tipo[Y] das criaturas de Brooks	62
3.1.2 As criaturas de Brooks e o experimento do bastão	65
3.2 Pengi: uma implementação da teoria da atividade.....	69
3.2.1 A RM do tipo[Y] em Pengi.....	79
3.2.2 Pengi e o experimento do bastão	85
Capítulo 4 – Conclusão.....	88
4.1 Apreciação da proposição «P2» e dos seus objetivos.....	89
4.2 Apreciação da proposição «P1».....	92
4.3 Semelhanças entre os Tipos de RM [Y] e a RM em Hayy	92

4.4 Implicações em outras áreas dos resultados apurados	93
Bibliografia.....	97
Fontes primárias.....	97
Fontes secundárias	99
Anexos	102
Anexo 1 – Diagrama conceitual das ciências cognitivas	103
Anexo 2 – Arquitetura de Von Neumann	104
Anexo 3 – Modelo esquemático de um sistema computacional.....	105
Anexo 4 – Exemplo de classificação de um percetção	106
Anexo 5 – Mecanismo tetrocromático versus mecanismo tricromático.....	107
Anexo 6 – Diagrama genérico de um circuito combinacional.....	108

Declaração de honra

Declaro que a presente tese é de minha autoria e não foi utilizado previamente noutro curso ou unidade curricular, desta ou de outra instituição. As referências a outros autores (afirmações, ideias, pensamentos) respeitam escrupulosamente as regras da atribuição, e encontra-se devidamente indicadas no texto e nas referências bibliográficas, de acordo com as normas de referenciação. Tenho consciência de que a prática de plágio e autoplágio constitui um ilícito académico.

Porto, 09 de outubro de 2019

Romeu Ivoleta Neto

Agradecimentos

Primeiro, eu agradeço ao meu orientador Professor Dr. João Alberto Pinto pela paciência e valorosa contribuição em todas as etapas da produção desta tese. Agradeço também a minha companheira, Beatriz Ferri Conejo, por me apoiar em cada passo da minha jornada neste mestrado. Deixo igualmente meus calorosos agradecimentos aos demais professores do Departamento de Filosofia da Universidade do Porto, pelos inesquecíveis momentos de convivência e partilha de conhecimento nos seminários de filosofia. Por fim, e não menos importante, agradeço a dois grandes filósofos que marcaram profundamente a minha alma: Professor Dr. Clóvis de Barros Filho e Professora Lúcia Helena Galvão. A Professora Lúcia Helena Galvão foi a responsável por me apresentar a magnífica filosofia oriental, já o Professor Clóvis de Barros Filho, foi aquele que me incitou o amor pelo conhecimento, e principalmente, despertou em mim a “vida que vale a pena ser vivida”.

Resumo

As últimas décadas têm sido marcadas por profundas mudanças, resultantes do acelerado progresso e alastramento da tecnologia em todas as dimensões da sociedade humana. Como corolário, torna-se crescente a percepção de que a cognição humana será superada pela inteligência artificial (IA) em algum ponto deste processo. Contudo, tal superação só é possível se a IA conseguir portar e operacionalizar a mesma estrutura de representação mental (RM) utilizada pela mente humana, e desta forma, incrementar sua cognição de forma autônoma. Em vista disso, a tese irá investigar acerca da estrutura adequada de RM necessária para a autonomia cognitiva em sistemas de IA, através do contraste entre o livro *O Filósofo Autodidata*, clássico da filosofia árabe acerca da autonomia cognitiva, com textos contemporâneos e de cunho técnico, que negam a necessidade da RM no processo cognitivo em sistemas de IA (autores antirrepresentacionistas). Desse modo, a tese irá demonstrar que esses autores não excluem a RM de fato, mas sim defendem outro[s] tipo[s] de estrutura[s] de RM, que chamaremos de tipo ‘y’. Resta saber, portanto, se a[s] estrutura[s] de RM do tipo ‘y’ atendem as necessidades da autonomia cognitiva na IA a mesma proeza cognitiva, e sob as mesmas circunstâncias, alcançadas por Hayy – o protagonista de *O Filósofo Autodidata*.

Palavras-chave: inteligência artificial, cognição, representação mental.

Abstract

The last decades have been marked by profound changes, as consequence of the rapid progress and dissemination of technology in all dimensions of human society. As a result, the perception that human cognition will be overcome by artificial intelligence (AI) at some point in this process grows. However, such overcoming is only possible if AI proves to be able to support and operate the same structure of mental representation (MR) present in the human mind, and thus increase its cognition autonomously. Therefore, the thesis will investigate the structure of MR that is necessary for cognitive autonomy in AI systems, through the contrast between the book *The Self-Taught Philosopher*, the classic of the Arab philosophy on cognitive autonomy, and contemporary and technical texts, which deny the necessity of MR in the cognitive process in AI systems (antirepresentationalists authors). Thereby, the thesis will demonstrate that these authors do not deny the MR, but instead of that, they defend other[s] types[s] of MR structure[s], that we will call 'y' type. Thereafter, it is necessary to know if the structure[s] of MR of 'y' type meets the need for cognitive autonomy of the AI systems and in circumstances similar to those imposed on Hayy – the protagonist from *The Self-Taught Philosopher* book – the same cognitive feat achieved by him.

Keywords: artificial intelligence; cognition; mental representation.

Índice de ilustrações

Figura 1 – Ilustração de uma rede neural simples	32
Figura 2 – Total de menções em papers de reinforcement learning comparado a outras técnicas de machine learning	35
Figura 3 – Gráfico de frequência de palavra a cada 1000 em papers entre os anos 2000 e 2018	36
Figura 4 – Suposto modelo cognitivo de Hayy, baseado na primeira fase de sua vida – 0 aos 7 anos.....	47
Figura 5 – Decomposição funcional	59
Figura 6 – Behavior-based decomposition.	61
Figura 7 – Topologia das máquinas de estados finitas (MEF) dos robôs de Brooks	62
Figura 8 – MEF de um torniquete	64
Figura 9 – Tela do jogo Pengo.....	71
Figura 10 – Finding the-block-to-kick-at-the-bee when lurking behind a wall.	77
Figura 11 – Modelo da teoria da atividade em Pengi	81
Figura 12 – PVR processando o aspecto the-block-that-the-block-I-just-kicked-will-collide-with.....	84

Lista de abreviaturas e siglas

CPU – unidade central de processamento

ET – etapa

IA – inteligência artificial

MT – máquina de Turing

MEF – máquina de estados finita

TCCM – teoria da computação clássica da mente

PVR – processador visual de rotinas

RM – representação mental

S – estado

Introdução

No século XII o filósofo Ibn Tufayl escreveu o romance *O Filósofo Autodidata*¹. A história conta a saga do protagonista Hayy Ibn Yaqzan que cresce sozinho em uma ilha deserta desde a mais tenra idade. Totalmente isolado e sem nenhum contato com outros humanos, Hayy no transcorrer dos anos vai desenvolvendo um contínuo processo de aprendizado, decorrente dos acontecimentos corriqueiros da ilha. Sem ninguém para conduzi-lo em sua jornada cognitiva, todo o seu aprendizado parece derivar da sua experiência. Após diversos acontecimentos, ao atingir a idade de 49 anos o personagem atinge um elevado grau de sabedoria. Contudo, como é possível alcançar isso naquelas circunstâncias?

O Filósofo Autodidata pode ser também apreciado como uma fonte de pesquisa para a área da inteligência artificial (IA), permitindo refletir sobre a cognição, e mais especificamente sobre a autonomia cognitiva, a partir das lentes de um clássico da filosofia árabe do século XII. Considerando o enredo do livro, dois importantes elementos podem ser apontados para a autonomia cognitiva: o primeiro elemento é a motivação provocada pelas adversidades impostas pelo ambiente hostil da ilha; o segundo elemento é a capacidade de transpor tais adversidades através da operacionalização de representações mentais (RM), possibilitando um incremento cognitivo contínuo de Hayy.²

Hipoteticamente, as RM funcionam como símbolos que representam o mundo na

¹ Ibn Tufayl, *O Filósofo Autodidata*, trad. Isabel Loureiro, Edição: 1 (São Paulo: UNESP, 2005).

² O *Filósofo Autodidata*, por causa do seu estilo literário, não deve ser considerado como um mero romance, mas como um texto filosófico – discussão semelhante acontece acerca de algumas obras de Friedrich Nietzsche. O professor de filosofia da Universidade de Vanderbilt, Lenn E. Goodman, aponta três dimensões distintas presentes nesse livro: educacional, religiosa e social – Ibn Tufayl, *Hayy Ibn Yaqzan*, ed. e trad. por Leen Evan Goodman, 5ª ed (Los Angeles: Gee tee bee, 2003). Vale ressaltar que a proeza cognitiva de Hayy é uma tentativa de Tufayl explicar a teoria da emanção, concebida pelo filósofo grego Plotino – tema não abordado nesta tese. Outro ponto importante relacionado a obra de Tufayl é aquele referente a possibilidade de uma criança ser criada por animais e conseguir sobreviver. A literatura científica narra poucos casos verídicos de crianças que viveram essa experiência. Dentre os mais notáveis, está o caso Oxana Malaya, que foi criada entre cachorros dos três aos oito anos de idade, tendo como consequência graves sequelas cognitivas, principalmente relacionadas à linguagem. “Oxana Malaya”, in Wikipédia, a enciclopédia livre, 11 de maio de 2019, https://pt.wikipedia.org/w/index.php?title=Oxana_Malaya&oldid=55110507.

mente de um agente. Tais símbolos têm características semânticas, referenciais e condições de utilização, mais ou menos bem-sucedida. Deste modo, a noção geral de representação pode ser considerada a noção mais básica – e uma das mais problemáticas – relacionada as ciências cognitivas e consequentemente a autonomia cognitiva em sistemas de IA. Se a RM é uma condição necessária para a cognição humana, a IA deverá dispor desse recurso para, no mínimo, igualá-la – o que não acontece até o presente momento.

Voltando a obra de Tufayl, como se pode supor que Hayy dispunha de representações mentais na trama? A partir do enredo da obra, o autor aponta evidências para este fato, nomeadamente quando diz que as experiências «permaneciam» na mente do protagonista.³ Desse modo, as RM estariam presentes nas experiências positivas e negativas de Hayy, como por exemplo, quando os animais batiam nele na disputa por alimentos e a sua frustração por não possuir as mesmas características dos demais animais (chifres, pelos ou garras).⁴

Com relação a cognição do protagonista, pode-se supor – e nesta introdução superficialmente – que o modelo cognitivo de Hayy seja fundamentado no suporte e na operacionalidade de RM – com criação de RM, associação entre RM e atualização de RM – gerando-se assim um ciclo incremental cognitivo autônomo indeterminado. E aqui se entenda «indeterminado» porque não é possível determinar o limite de capacidade cognitiva. Isto posto, a tese traz como objetivo secundário, identificar o tipo de estrutura de RM utilizada por Hayy e tentar verificar se tal configuração também é uma condição necessária para a autonomia cognitiva em sistemas de IA.

Com relação ao objetivo principal, a tese investigará a relação entre tipo[s] de RM e autonomia cognitiva em sistemas de IA, mediante textos de cunho técnico e

³ « Com o tempo, passou a se lembrar das coisas mesmo quando elas estavam ausentes; percebeu que umas lhe inspiravam desejo, outras aversão. » Ibn Tufayl, *O Filósofo Autodidata*, trad. Isabel Loureiro, Edição: 1 (São Paulo: UNESP, 2005), 63.

⁴ « Enquanto isso, observava todos os animais que via cobertos de pelos lanosos ou sedosos, ou de plumas. Observava sua rapidez na corrida, sua força, as armas de que haviam munido para lutar contra o adversário, tais como chifres, presas, cascos, esporões, garras. Voltado a si mesmo, via-se nu e sem armas, lento na corrida fraco contra os animais que lhe disputavam os frutos, se apropriavam deles em seu detrimento e lhos tiravam sem que pudesse enxotá-los ou escapar-lhes. » Ibn Tufayl, *O Filósofo Autodidata*, trad. Isabel Loureiro, Edição: 1 (São Paulo: UNESP, 2005), 63.

antirrepresentacionista. Esses textos descrevem as implementações dessas ideias por meio de experimentos técnicos: robótica em *Intelligence Without Representation*⁵ e sistêmica em *Pengi: An Implementation of a Theory of Activity*⁶. Apesar de ambos textos tutelarem que a RM não é uma condição necessária na cognição (antirrepresentacionalismo), a tese irá provar o contrário, que de fato há outros tipos de RM em curso nessas propostas.

Por fim, resta saber se os tipos de RM antirrepresentacionistas são capazes de atingir o mesmo tipo de proeza cognitiva conquistada por Hayy, e principalmente, sob as mesmas circunstâncias impostas a ele.

Observações sobre palavras-chave

Com relação as palavras-chave que permearão a tese, faz-se necessário explicitar o seu significado, para que a averiguação possa proceder sem equívocos interpretativos. A primeira é a palavra-chave «cognição», derivada do termo latim *cognoscere*, que abrange como alguns de seus significados⁷:

- I. aprender,
- II. descobrir,
- III. ter experiência de,
- IV. conhecer pela experiência,
- V. tornar-se ciente de,
- VI. discernir,
- VII. reconhecer (uma pessoa ou alguma coisa já conhecida).

Deste modo, para algo ser considerado agente «cognoscente», ou em outros termos, capaz de «cognição», a agência deverá revelar todas as condições supracitadas e de forma «autônoma». E eis aqui outra palavra-chave que será importante no contexto desta tese. Juntamente com «cognoscente», «autônomo» define o agente como capaz de

⁵ Rodney A. Brooks, “Intelligence Without Representation”, *Artificial Intelligence* 47, nº 1–3 (janeiro de 1991): 139–59, [https://doi.org/10.1016/0004-3702\(91\)90053-M](https://doi.org/10.1016/0004-3702(91)90053-M).

⁶ Philip E. Agre e David Chapman, “Pengi: An Implementation of a Theory of Activity”, in *Proceedings of the Sixth National Conference on Artificial Intelligence - Volume 1*, AAAI’87 (AAAI Press, 1987), 268–272, <http://dl.acm.org/citation.cfm?id=1863696.1863744>.

⁷ “Cognoscere”, in *Oxford Latin Dictionary* (Oxford: Oxford University Press, 1968), 346.

realizar e autogovernar a cognição de forma independente.

Além dessas palavras-chave, outras duas serão necessárias no desenvolvimento da tese: «estrutura» e «tipo» de RM. Começando pelo «tipo», termo que será usado para identificar conjuntos de RM, organizadas de acordo com o específico programa de investigação que a elas crucialmente se referem⁸: (1) RM de tipo simbólico no cognitivismo; (2) RM de tipo subsimbólico no connexionismo; e (3) RM de tipo enacionado no enativismo. Contudo, esta lista pode precisar de ser estendida se ao longo da tese novos tipos de RM forem identificados.

A «estrutura», por sua vez, corresponde ao modo como cada RM é pensada dentro de cada «tipo». Isto significa que cada «tipo» de RM está ligado a uma perspectiva diferente sobre a construção, organização e apresentação das RM. No primeiro capítulo deste trabalho serão apresentados mais detalhadamente os tipos de RM, assim como a especificação das respectivas estruturas.

Metodologia, objetivos e estrutura

Voltando ao tema central da tese, o objetivo principal será investigar qual tipo de RM é necessário para a autonomia cognitiva em sistemas de IA. Os tipos serão nomeados provisoriamente [X]. Desta forma, a ideia principal da tese consiste em investigar a seguinte questão:

Que relação há entre sistemas de IA possuírem estrutura[s] de RM do[s] tipo[s] [X] e tais sistemas apresentarem autonomia cognitiva?

Mas como identificar o[s] tipo[s] [X] de RM? A questão acima permite identificar duas proposições básicas:

p: Sistemas de IA apresentam autonomia cognitiva.

q: Sistemas de IA possuem estrutura[s] de RM do[s] tipo[s] [X].

Estas duas proposições, por sua vez, podem estar relacionadas dos seguintes dois modos:

PI – Possuir estrutura de RM do tipo [X] é condição necessária para a autonomia cognitiva em sistemas de IA. ($p \rightarrow q$)

⁸ Ver Anexo 1 – Diagrama conceitual das ciências cognitivas.

P2 – Sistemas de IA apresentam autonomia cognitiva, mas não possuem estrutura[s] de RM do[s] tipo[s] [X]. ($p \wedge \sim q$)

À vista disso, a tese assume que a forma mais eficiente de encontrar a RM do tipo [X] é averiguando aqueles programas que fazem objeção a necessidade da RM para a cognição em sistemas de IA. Consequentemente, tais programas devem oferecer propostas capazes de explicar a cognição sem o artifício da RM. Contudo, esta tese assume a posição de que os programas que não fazem uso da RM no processo cognitivo, ou seja, os programas antirrepresentacionistas, utilizam-se inevitavelmente de algum tipo de RM, identificados provisoriamente como sendo de tipo [Y]. Isto posto, resta averiguar se as[s] RM[s] do[s] tipo[s] [Y], que um pouco paradoxalmente se encontram nos programas antirrepresentacionistas, atendem a necessidade da autonomia cognitiva em sistemas de IA, e se tais tipos se assemelham de alguma forma como o tipo de RM identificado em Hayy.

Dessa forma, a averiguação do tipo necessário de RM para a autonomia cognitiva em sistemas de IA, acontecerá através da investigação sobre «P2». Caso a proposição seja verdadeira, o argumento principal desta tese torna-se inválido pelo princípio da contradição com a «P1». Caso a «P2» seja falsa, os seguintes objetivos serão alcançados:

Objetivo 1: A tese irá mostrar que os programas antirrepresentacionistas usam RM do[s] tipo[s] [Y] e, portanto, também são representacionistas, diferenciados apenas pelos tipos de RM.

Objetivo 2: A tese irá investigar se as RM do[s] tipo[s] [Y] identificadas nos textos, satisfazem a necessidade de autonomia cognitiva em sistemas de IA, através do experimento elaborado a partir de um excerto do livro de Tufayl, denominado «o experimento do bastão».⁹

Objetivo 3: A tese irá apontar através dos resultados obtidos nos objetivos 1 e 2, quais RM do[s] tipo[s] [Y] são necessárias para a autonomia cognitiva em sistemas de IA, e desta forma, classificá-las como RM do[s] tipo[s] [X].

Objetivo 4: A tese irá apontar através dos resultados obtidos nos objetivos 1 e 2,

⁹ Cfr. seção 2.2 O experimento do bastão, 46.

quais RM do[s] tipo[s] [Y] não são necessárias para a autonomia cognitiva em sistemas de IA e desta forma, excluí-las como RM do[s] tipo[s] [X].

Vale ressaltar que o objetivo principal da tese não é investigar o modelo cognitivo presente em Hayy, mas sim investigar o tipo necessário de RM necessário para a autonomia cognitiva em sistemas de IA. Assim, o exemplo de Hayy será usado apenas como referência comparativa, possibilitando averiguar o problema relacionado a cognição de forma peculiar e distinta daquelas já averiguadas até o momento.

Por fim, a tese apresentará a estrutura de desenvolvimento a seguir esboçada.

No primeiro capítulo serão apresentadas as definições de RM, seus tipos e estruturas, a partir dos principais programas definidos pela ciência cognitiva: cognitivismo, connexionismo e enativismo. Após esta breve apresentação, serão definidas as teses cruciais para cada tipo específico, a fim de que possam ser devidamente averiguadas e comparadas no decorrer da tese.

No segundo capítulo será averiguado qual modelo cognitivo, segundo as definições dos tipos de RM propostos pela ciência cognitiva e a trama de Ibn Tufayl, melhor explicam a autonomia cognitiva em Hayy. Este modelo será usado como referência ao longo da tese no contraste com as RM do tipo [Y].

No terceiro capítulo serão averiguados textos que defendem modelos cognitivos e de inteligência sem o uso da RM e de cunho técnico.

O primeiro texto analisado será *Intelligence Without Representation*¹⁰ de autoria do filósofo Rodney Brooks. Neste texto o autor defende um modelo incremental de inteligência sem o uso da RM. Brooks expõe que o sistema de inteligência não é formado por um sistema central de representação que interage com outros sistemas centrais. Pelo contrário, a inteligência seria decomposta em sistemas paralelos e independentes que interagem com o mundo através da percepção e da ação.

The fundamental decomposition of the intelligent system is not into independent information processing units which must interface with each other via representations.

¹⁰ Brooks, “Intelligence Without Representation”.

Instead, the intelligent system is decomposed into independent and parallel activity producers which all interface directly to the world through perception and action [...].¹¹

O texto seguinte *Pengi: An Implementation of a Theory of Activity*¹², de Philip E. Agre e David Chapman apresenta uma implementação técnica de um sistema dinâmico cognitivo, sem a configuração prévia de modelos do mundo e regras. *Pengi* foi construído a partir de rotinas, entidades e aspectos, não constituindo um modelo representacional da mente.

Action arbitration has many of the benefits of planning, but is much more efficient, because it does not require representation and search of future worlds. In particular, a designer who understands the game's common patterns of interaction (its "dynamics") can use action arbitration to produce action sequencing, nonlinear look ahead to resolve goal interactions, and hierarchical action selection.¹³

No capítulo final serão apreciadas as análises dos autores estudados, bem como a resposta da pergunta que rege a tese: qual o tipo de RM é necessária para a autonomia cognitiva em sistemas de IA?

Uma resposta à pergunta anterior pode servir para corroborar ou refutar teses envolvidas em discussões, pesquisas e movimentos relacionados à inteligência artificial, como, por exemplo, a *Open Letter on Artificial Intelligence*¹⁴. Este documento foi assinado por personalidades altamente reconhecidas no meio acadêmico e empresarial, tendo como alguns dos seus signatários Elon Musk e Stephen Hawking. O objetivo desta carta é alertar a sociedade dos perigos que podem advir da IA. Dentre os perigos apontados está a possibilidade de a racionalidade artificial superar a humana e colocar em risco o bem-estar e a própria existência da humanidade.

Esta preocupação pode ser confirmada através do principal projeto vigente da

¹¹ Brooks, 139.

¹² Philip E. Agre e David Chapman, "Pengi: An implementation of a Theory of Activity", in *Proceedings of the Sixth National Conference on Artificial Intelligence - Volume 1*, AAAI'87 (AAAI Press, 1987), 268–72, <http://dl.acm.org/citation.cfm?id=1863696.1863744>.

¹³ Agre e Chapman, 271–72.

¹⁴ "Open Letter on Artificial Intelligence", in *Wikipedia*, 4 de abril de 2018, https://en.wikipedia.org/w/index.php?title=Open_Letter_on_Artificial_Intelligence&oldid=834163483.

empresa Google. No livro *Mundo sem mente: a ameaça existencial da alta tecnologia*¹⁵, o autor Franklin Foer denuncia a obsessão ideológica e técnica de um dos fundadores do Google, Larry Page, de executar e implementar a IA em escala global.¹⁶ O projeto em questão chama-se *Google Brain*¹⁷ e tem como diretor de engenharia nada menos que o renomado cientista Ray Kurzweil, idealizador da teoria da singularidade. Esta teoria defende a ideia de que, muito em breve, a IA terá a capacidade de se autorreplicar e autoincrementar de forma exponencial, transformando-se numa superinteligência que impactará profundamente e de forma irreversível a civilização humana.

A singularidade refere-se a uma ruptura no contínuo espaço-temporal – descreve o momento em que o finito se torna infinito. Para Kurzweil, a singularidade será quando a inteligência artificial se tornar todo-poderosa, quando os computadores forem capazes de conceber e de construir computadores. Obviamente, tal superinteligência vai criar uma superinteligência ainda mais poderosa – e assim sucessivamente ao longo das gerações pós-humanas. Nesse momento, tudo poderá acontecer – uma IA e uma nanotecnologia fortes poderão criar qualquer produto, qualquer situação, qualquer ambiente que imaginemos.¹⁸

Independentemente das conclusões que serão apresentadas, poderão gerar-se reflexões adicionais sobre os limites da inteligência artificial. Se a representação mental é uma condição necessária ao processo cognitivo e não pode ser reproduzido pela IA, então a sociedade deverá ponderar quais atividades devem ser entregues a este tipo de tecnologia – pois neste caso ela sempre será uma redução da inteligência humana.

Porém, se a representação mental não for uma condição necessária à autonomia cognitiva da IA, e for possível tecnicamente atingir tal autonomia, então será uma questão de tempo até enfrentarmos uma tecnologia que não somente superará a

¹⁵ Franklin Foer, *Mundo sem mente: a ameaça existencial da alta tecnologia*, trad. Luís Oliveira Santos, 1º ed (Temas e Debates – Círculo de Leitores, 2018).

¹⁶ « A IA do Google é de tal modo eficaz que até nos fornece os resultados da pesquisa antes de acabarmos de escrever. Todavia, como herdeiro da elevada tradição da IA, para Larry Page, este feito é um passo insignificante no caminho para a concretização de uma missão bastante mais profunda – uma missão em ambos os sentidos do termo, religioso e científico. Page criou a empresa para alcançar aquilo que se chama “IA completa”, a criação de máquinas com a capacidade para igualar, e a seu tempo ultrapassar a inteligência humana. » Foer, 50.

¹⁷ Foer, 65.

¹⁸ Foer, 59–60.

capacidade cognitiva humana, como também a questionará e a confrontará, além das inevitáveis e impensáveis implicações nas áreas da ética, política, economia ou mesmo da religião.

Por fim, esta tese tem como meta contribuir filosoficamente para os debates referentes aos limites da capacidade cognitiva da inteligência artificial – superar a capacidade humana ou ser uma ferramenta poderosa de auxílio ao homem, mas sem a capacidade cognitiva de ultrapassá-lo.

Capítulo 1 – A representação mental e seus tipos

Neste capítulo será apresentado o conceito de representação mental, assim como os tipos e as estruturas de RM, a partir dos programas cognitivista, conexionista e enativista.

1.0 Representação mental «RM»

A RM é hipoteticamente a representação do mundo na mente de um agente. Citando como exemplo uma maçã, ela pode ser representada de diferentes maneiras: pela palavra *maçã*; pela palavra *apple*; pela pintura da maçã ou até mesmo por um desenho estilizado. Todas estas coisas podem tornar-se complexos e robustos objetos cognitivos uma vez articulados com o sabor ácido, a cor vermelha, a forma arredondada, referências à história do pecado original, às descobertas de Isaac Newton, à expulsão de Adão e Eva do paraíso e à marca da empresa de tecnologia mais valiosa do mundo (Apple). As RM podem fazer parte de estados cognitivos internos dos agentes, uma vez que as propriedades cognitivas pressupõem o aprendizado, a associação e o fortalecimento de propriedades semânticas via percepção e/ou raciocínio.

The notion of a “mental representation” is, arguably, in the first instance a theoretical construct of cognitive science.¹⁹

Isto posto, as RM assumem como principal característica a sua intencionalidade ou direção para objetos encenados na mente de um agente.

À vista disto, a definição de RM nesta tese seguirá as seguintes condições:

- I. RM possuem intencionalidade, são tipicamente direcionadas para objetos;
- II. RM são encenações do mundo na mente de um agente;
- III. RM fazem parte de estados cognitivos internos.

Ao longo desta tese será averiguada a necessidade das RM para a autonomia

¹⁹ David Pitt, “Mental Representation”, org. Edward N Zalta, Uri Nodelman, e Colin Allen, *The Stanford Encyclopedia of Philosophy*, Winter Edition de 2018, 1.

cognitiva em sistemas de IA. Desse modo, resta saber se os principais programas cognitivos, alguns deles antirrepresentacionistas, conseguem atender a autonomia cognitiva, e principalmente, sob as mesmas adversidades impostas ao Hayy na trama de Tufayl. Tais questões assim como as estruturas envolvidas nestes programas cognitivos serão apreciados no decorrer dos próximos capítulos.

1.1 Tipo 1 – a representação mental simbólica (cognitivismo)

A ciência cognitiva começou a ser delineada a partir de 1943, propondo-se a estudar a mente sob lentes distintas daquelas propostas até então pela filosofia e psicologia. Consequentemente, esta abordagem expandiu as perspectivas de pesquisa e compreensão da mente, graças às contribuições fornecidas pelas novas disciplinas constituintes à ciência cognitiva: matemática, biologia, neurociência e a incipiente computação. Esta fase durou até o ano de 1953 e durante este período, cientistas como Alan Turing²⁰, John von Neumann²¹, Warren McCulloch e Walter Pitts²² criaram teorias que fundamentaram e foram vitais para o desenvolvimento dos computadores digitais e os sistemas de IA, juntamente com a disseminação e consolidação da teoria que viria a influenciar as ciências cognitivas: a teoria computacional clássica da mente (TCCM). Estes fatos imprimiram um caráter cibernético – ou, mais em geral, tecnológico – à ciência cognitiva logo desde o seu início.

²⁰ A contribuição de Alan Turing para a área da ciência da computação e IA é incalculável. No artigo de Alan Turing, “On computable numbers, with an application to the Entscheidungsproblem”, *Proc. London Maths. Soc.*, 2, 42 (1936): 230–65, o autor definiu a máquina de Turing, que é uma concepção abstrata de um computador com ilimitado tempo e espaço, capaz de computar símbolos. O segundo trabalho de grande relevância foi outro texto de Alan Turing, “Computing machinery and intelligence”, *Mind* 50 (1950): 433–60, que apresenta o «Teste de Turing» cuja finalidade é verificar através de um experimento, se um interlocutor não visto, poderia responder questões de tal forma que dificultasse a sua identificação como máquina ou humano – neste cenário o computador teria passado no teste.

²¹ John von Neumann criou um modelo de arquitetura – Arquitetura de von Neumann – que permite as máquinas digitais armazenarem os seus programas no mesmo espaço de memória de seus dados. A arquitetura é composta de memória, unidade de controle, unidade aritmética/lógica e entrada/saída de dados. As arquiteturas dos computadores atuais são baseadas neste modelo. Ver Anexo 2 – Arquitetura de von Neumann.

²² No clássico artigo: Warren S. McCulloch e Walter H. Pitts, “A Logical Calculus of Ideas Immanent in Nervous Activity”, *Bulletin of Mathematical Biophysics* 5 (1943): 115–33, os autores defendem a teoria de que a lógica seria a disciplina adequada para a compreensão das atividades do cérebro. Além desta ideia é apresentada a tese de que o cérebro seria um dispositivo que incorpora princípios lógicos em seus elementos constituintes – os neurônios. Estes elementos teriam dois estados – ativo e inativo – permitindo uma abrangente estrutura lógica.

A ciência cognitiva entrou num novo rumo a partir do ano de 1956, quando cientistas como Herbert Simon, Noam Chomsky, Marvin Minsky e John McCarthy, ainda que influenciados pela fase cognitiva predecessora – a fase cibernética – mas insatisfeitos com a abordagem behaviorista²³ para o estudo do comportamento humano, criam uma nova perspectiva para o entendimento da mente, denominada normalmente «cognitivismo».

A partir daqui o cognitivismo se estabelece como o programa de investigação preeminente, pleiteando um modelo mental no qual a inteligência e a cognição se assemelham a dinâmica funcional de um sistema computacional – que constitui a influência direta da TCCM. A mente teria capacidade de computar símbolos – os seus atributos físicos, e não propriamente a semântica, de modo semelhante ao processo realizado nos sistemas computacionais. Assim, a cognição e a inteligência neste programa seriam explicadas pela capacidade do agente portar e realizar uma computação simbólica.

Contudo, a comparação da mente com um sistema computacional pode gerar muitas interpretações confusas que será-o apontadas nesta tese. Dentre os equívocos está aquele relacionado a associação literal a «metáfora do computador» ao invés de um «sistema computacional»²⁴. Este discernimento terminológico faz-se necessário, para evitar enganos frequentes, como por exemplo, as objeções direcionadas ao cognitivismo referentes a propriedade não programável da mente. Apesar disso, esta objeção não inviabiliza a proposta cognitivista em agentes biológicos porque a maioria das máquinas

²³ A teoria behaviorista defende a tese de que a mente é uma entidade falível e duvidosa para o estudo dos comportamentos e estes devem, portanto, serem estudados e mapeados através de métodos científicos e quaisquer referências a estados subjetivos devem ser excluídas. Em contrapartida, o cognitivismo defende que os pensamentos influenciam os comportamentos e não podem ser excluídos da investigação comportamental dos agentes.

²⁴ Um computador é a máquina capaz de computar/tratar informações de vários tipos de forma automática. Todo computador possui um sistema computacional, composto de dispositivos eletrônicos (hardwares) que processam informações através de programas (softwares). Um sistema computacional é, de facto, formado pelos seguintes componentes: hardware, sistema operacional, programas de aplicação e usuários. (Ver Anexo 3 – Modelo esquemático de um sistema computacional). Outra nota de atenção é que se a proposta é literalmente explicar o funcionamento da mente sob as lentes de sistemas computacionais, os componentes destes sistemas poderão tanto ajudar quanto complicar a explicação da dinâmica mental. Por exemplo, o sistema operacional – responsável por orquestrar operações de hardware, software e usuário – poderia aparecer como uma objeção ao cognitivismo ou demandar explicações sobre como ocorre a coordenação de todo o sistema mental.

de Turing – a referência computacional do cognitivismo – não são programáveis.²⁵

Thus, the phrase “computer metaphor” strongly suggests theoretical commitments that are inessential to CCTM [TCCM]. The point here is not just terminological. Critics of CCTM [TCCM] often object that the mind is not a programmable general purpose computer [...]. Since classical computationalists need not claim (and usually do not claim) that the mind is a programmable general purpose computer, the objection is misdirected.²⁶

O segundo ponto a ser esclarecido é considerar «metaforicamente» a mente como um sistema computacional e não de forma «literal». A TCCM afirma «literalmente» que a mente é um sistema computacional e não semelhante a um. Contudo, o modelo não é restritivo às estruturas artificiais e pode ser operacionalizado em outros tipos de estruturas, inclusive as biológicas, uma vez que se trata de um modelo abstrato.

We attain an abstract computational description that could be physically implemented in diverse ways (e.g., through silicon chips, or neurons, or pulleys and levers). CCTM [TCCM] holds that a suitable abstract computational model offers a literally true description of core mental processes.²⁷

O terceiro ponto a ser esclarecido e não menos importante é aquele relacionado a máquina de Turing (MT) – ou ao slogan «*the mind is a Turing machine*», muitas vezes também usado de forma errônea. O modelo computacional de Turing foi incorporado no programa cognitivista, porém, o formalismo que rege este modelo é intratável quando se trata dos agentes biológicos e precisa ser redirecionado apropriadamente em cada caso

²⁵ Aqui é preciso uma nota de atenção: apesar do programa cognitivista não afirmar que uma mente precisa ser programável, isto não trata da questão relacionada a sintaxe–semântica dos símbolos. No caso dos sistemas de IA, as sintaxes garantem o significado dos símbolos – definido pelos programadores. Porém no caso dos agentes biológicos não é claro como isto ocorre. « *As we mentioned in a computer program the syntax of the symbolic code mirrors or encodes its semantics. In the case of human language, it is far from obvious that all of the semantic distinctions relevant in an explanation of behavior can be mirrored syntactically. [...] Furthermore, although we know where the semantic level of a computer’s computations comes from (the programmers), we have no idea how the symbolic expressions supposed by the cognitivist to be encoded in the brain would get their meaning.* » Francisco J. Varela, Evan Thompson, e Eleanor Rosch, “Symbols: the cognitivist hypothesis”, in *The Embodied Mind: Cognitive Science and Human Experience*, revised edition (Cambridge, Massachusetts ; London England: MIT Press, 2016), 42.

²⁶ Michael Rescorla, “The Computational Theory of Mind”, org. Edward N Zalta, Uri Nodelman, e Colin Allen, 2017, 8–9.

²⁷ Rescorla, 9.

específico.

Como exemplo pode-se citar duas características típicas da MT que são restritivas para agentes biológicos e que precisam ser adaptadas: a dinâmica da entrada/saída dos símbolos e a capacidade infinita de memória. Esta última característica deverá ser adequada para um modelo biológico com capacidade de armazenamento abrangente, porém finita. Por sua vez, a dinâmica da entrada/saída dos símbolos na MT é definida em alocações de memória, em contrapartida na agência biológica dá-se por sistemas de percepção. Logo, faz-se necessário uma teoria que descreva como a computação mental realiza esta interface das entradas dos estímulos sensoriais com as respectivas saídas motoras.²⁸

CCTM [TCCM] claims that mental activity is “Turing-style computation”, allowing these and other departures from Turing’s own formalism.²⁹

Em relação a cognição, no cognitivismo este processo é operacionalizado pela manipulação das propriedades físicas do símbolo, sendo as propriedades semânticas definidas sintaticamente – ressaltando que a articulação sintática-semântica é compreensível em sistema de IA, mas um tanto quanto incerta nos agentes biológicos³⁰. De qualquer maneira, a cognição torna-se possível em qualquer agente capaz de portar e operacionalizar símbolos discretos – elementos distintos e desconexos entre si – e um símbolo pode ter N representações físicas. Por consequência, o mesmo valor semântico também pode ser obtido com diferentes símbolos – imagens, palavras ou outras formas quaisquer.

Tendo como exemplo a relação operacional entre uma RM e suas propriedades semânticas, pode-se supor que um símbolo para maçã seja uma RM que indexa um

²⁸ Rescorla, 9.

²⁹ Rescorla, 10.

³⁰ Vale ressaltar novamente que os sistemas operam somente os atributos físicos dos símbolos e não a sua semântica. No caso dos sistemas de IA esta semântica é atribuída via sintaxe – num cenário que foi exemplificado no clássico experimento mental de John Searle, O Quarto Chinês. John R Searle, “Minds, Brains, and Programs”, *The Behavioral and Brain Sciences*, 1980, 42, <https://doi.org/10.1017/S0140525X00005756>. Neste texto o filósofo demonstra que as máquinas não conseguem atribuir significados às palavras, por mais que os humanos possam ter a percepção que sim. Já nos seres biológicos não há certeza sobre como este processo é operacionalizado e, portanto, esta é uma das questões que precisam ser respondidas pelo cognitivismo.

objeto do mundo – intencionalidade – dentre tantos outros representados na mente do agente. Porém, esta RM sozinha não explica o comportamento do ser, sendo necessário a relação com crenças (eu acredito que maçãs são saudáveis), desejos (eu desejo uma maçã) ou percepções (o gosto desta maçã é doce). Logo, a relação entre RM e estas coisas é que caracterizam as propriedades semânticas, e definem desta forma os estados mentais da agência.³¹

Em sistemas de IA as associações entre RM e conteúdos são definidas sintaticamente pelos programadores e o valor semântico de um objeto é o reflexo do que foi definido pela sintaxe – tabela de máquina^{32 33}. Hipoteticamente, pode-se supor que cada símbolo seja capaz de figurar numa máquina de estados, a qual especifica um determinado gênero de estado – crenças, desejos e percepções – cada qual passível de ser avaliado semanticamente. Em outros termos, para um símbolo «maçã» relacionado a um estado de desejo, «desejo comer uma maçã», haveria uma máquina de estados que possibilitaria atualizar o valor semântico deste símbolo.

Thus, representationally contentful mental states are *semantically evaluable* with respect to properties such as truth, accuracy, fulfillment, and so on. [...] Beliefs have truth-conditions (conditions under which they are true), perceptual states have accuracy-conditions (conditions under which they are accurate), and desires have fulfillment-conditions (conditions under which they are fulfilled).³⁴

De forma simples, os estados de tal máquina poderiam acabar por assumir a seguinte configuração:

S1. desejo de comer maçã não realizado;

³¹ « *Philosophers and cognitive scientists use the term “representation” in diverse ways. Within philosophy, the most dominant usage ties representation to intentionality, i.e., the “aboutness” of mental states. Contemporary philosophers usually elucidate intentionality by invoking representational content. A representational mental state has a content that represents the world as being a certain way, so we can ask whether the world is indeed that way.* » Michael Rescorla, “The Computational Theory of Mind”, org. Edward N Zalta, Uri Nodelman, e Colin Allen, 2017, 29–30.

³² Traduzido do termo inglês *machine table*.

³³ « *A machine table dictates which elementary operation the central processor performs, given its current machine state and the symbol it is currently accessing. The machine table also dictates how the central processor’s machine state changes given those same factors. Thus, the machine table enshrines a finite set of routine mechanical instructions governing computation.* » Michael Rescorla, “The Computational Theory of Mind”, org. Edward N Zalta, Uri Nodelman, e Colin Allen, 2017, 4.

³⁴ Rescorla, 30.

S2. desejo de comer maçã realizado.

Após o término de um fluxo da máquina de estados, as propriedades físicas do símbolo são computadas e o novo valor semântico do símbolo é atualizado: S2. desejo de comer maçã realizado. No final, a atribuição do valor semântico de um símbolo é um espelho³⁵ das regras implementadas na MT da agência. O importante é que a computação ocorre somente nas propriedades físicas do símbolo e não em sua semântica, que apenas reflete o que foi sintaticamente determinado na tabela de máquina.

A digital computer, however, operates only on the physical form of the symbols it computes; it has no access to their semantics value. Its operations are nonetheless semantically constrained because every semantic distinction relevant to its program has been encoded in the syntax of its symbolic language by the programmers. In a computer, that is, syntax mirrors or is parallel to the (ascribed) semantics.³⁶

Por fim, a RM do tipo [1], ou simbólica, será validada ao longo da tese somente se todas as condições determinadas nesta seção forem satisfeitas pela agência:

- I. O funcionamento da mente assemelha-se com uma máquina de Turing (MT);
- II. Os agentes devem ser capazes de possuir RM;
- III. A cognição é operacionalizada sintaticamente, através da computação das propriedades físicas e não das propriedades semânticas de símbolos;
- IV. A cognição é confirmada se após essa computação, a agência tiver êxito no problema/constrangimento imposto a ela.

1.2 Tipo 2 – a representação mental subsimbólica (conexionismo)

As principais ideias que viriam a fundamentar o conexionismo já estavam presentes nos primórdios do cognitivismo – aproximadamente a partir de 1956. As teorias que afirmavam a possibilidade de sistemas inteligentes operarem sem regras pré-

³⁵ Varela, Thompson, e Rosch, “The embodied mind”, 2016, 42.

³⁶ Varela, Thompson, e Rosch, 41.

estabelecidas e sem a necessidade de CPU³⁷ já eram notórias nesta fase. Dentre as teorias se destacam a do psicólogo Frank Rosenblatt, que em 1958 criou a primeira rede neural artificial nomeada de percetron³⁸, e a do neurologista William Ross Ashby, autor do primeiro estudo identificando padrões globais de comportamento em sistemas compostos de conexões aleatórias.³⁹

Contudo, estas ideias ficaram dormentes até a segunda metade da década de 70, quando a redescoberta em paralelo de teorias importantes na física (auto-organização de sistemas complexos), na matemática (sistemas não lineares) e o progresso da tecnologia (computadores mais velozes e acessíveis), deram o impulso para a retomada das ideias conexionistas como forma de contornar as adversidades e as limitações apresentadas pelo cognitivismo – o programa vigente e predominante até a segunda metade da década de 70.⁴⁰

Como o programa cognitivista segue as mesmas diretrizes de um sistema computacional, cria-se um vínculo no qual as deficiências apresentadas por este último, acabam inevitavelmente impactando aquele. Dentre os obstáculos, está o gargalo de *von Newman*, ou mais especificamente, o limite computacional desta arquitetura. Em razão deste modelo operar de forma linear e não em paralelo, cria-se um gargalo na capacidade máxima de processamento de instruções fazendo com que o sistema se torne

³⁷ CPU, abreviado do inglês *Central Processing Unit*, também conhecida como processador, tem a tarefa de realizar as instruções armazenadas – dados e programas – do computador. Este dispositivo é geralmente composto de uma memória principal, controle de unidade e unidade de aritmética-lógica (onde as operações lógicas são computadas). A CPU pode ser considerada como o cérebro do computador.

³⁸ O Percetron é um modelo cognitivo composto apenas de uma camada simples de componentes, quando há mais de uma camada é chamado de rede neural. Esse dispositivo é composto de 4 partes: entrada/saída de valores, pesos, somatório da rede e função de ativação, e tem como objetivo agir como um classificador binário. O objetivo de Rosenblatt foi demonstrar que as redes podiam ser treinadas e atuar na classificação de dados, como por exemplo mapear o total de pixels de uma imagem e classificá-la ao final do processo, como pertencente a categoria clara ou escura. Ver exemplo de classificação binária no perceptron: Anexo 4 – Exemplo de classificação de um percetron.

³⁹ Francisco J. Varela, Evan Thompson, e Eleanor Rosch, “Emergent properties and connectionism”, in *The Embodied Mind: Cognitive Science and Human Experience*, revised edition (Cambridge, Massachusetts ; London England: MIT Press, 2016), 85.

⁴⁰ Varela, Thompson, e Rosch, 85.

ineficaz⁴¹. Assim, operações mais exigentes, como por exemplo, o processamento de imagens e vídeos, acabam demandando mais recursos computacionais para serem levados a cabo. Consequentemente, novas requisições de processamento não poderão ser realizadas até que os processos anteriores sejam finalizados, o que inevitavelmente cria um obstáculo técnico. Outro problema apresentado por sistemas computacionais é aquele relacionado a contingência, pois caso ocorra um problema técnico num dispositivo, como por exemplo na CPU, ou mesmo em alguma parte interna dele, o sistema inteiro se torna inoperante e/ou instável. A vista disto, como o programa conexionista endereça tais questões?

Com relação ao gargalo de *von Newman*, o conexionismo defende a tese de que o cérebro está organizado de forma distribuída e não centralizada. Desse modo, a operação ocorre através da interconexão e auto-organização dos neurônios, o que elimina a necessidade de um dispositivo centralizador ou manipulador de símbolos e consequentemente, resolve o problema do gargalo de processamento. Como a operação é executada por N neurônios – processamento em paralelo –, se porventura alguns deles falharem na operação, a eficácia da operação não é comprometida, o que soluciona o problema relacionado a contingência operacional.

No que tange a estrutura das redes neurais, estas podem ser definidas como modelos simplificados do cérebro, compostos por miríades de unidades – análogos aos neurônios – e cujas forças de conexão – análogas às sinapses – variam de acordo com o peso e as regras de ativação consolidadas nos processos de aprendizagem. Além disso, são compostas por três classes: unidades de entrada, unidades ocultas e unidades de saída. Se uma rede neural for comparada ao sistema nervoso humano, as unidades de entrada seriam os neurônios sensoriais, as unidades de saída os neurônios motores e as unidades ocultas os demais neurônios.^{42 43}

⁴¹ Ressalta-se o problema relacionado às operações seriais e não em paralelo tem sido tratado pelos filósofos computacionalistas – aderentes a teoria TCCM. Por exemplo, no texto de Jerry Fodor e Zenon Pylyshyn, “Connectionism and cognitive architecture: a critical analysis”, *Cognition* 28 (1988): 3–71.

⁴² James Garson, “Connectionism”, org. Edward N Zalta, Uri Nodelman, e Colin Allen, *The Stanford Encyclopedia of Philosophy*, 2018, 2.

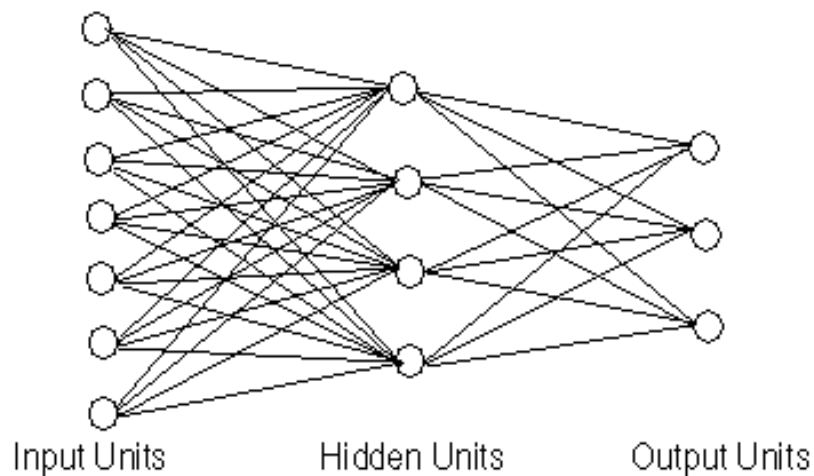


Figura 1 – Ilustração de uma rede neural simples

Em relação a dinâmica da cognição, no connexionismo não há necessidade de representações abstratas e computação de símbolos, conforme proposto no programa cognitivista. A cognição é operada pela miríade de neurônios que são acionados por regras particulares/locais, responsáveis pela emergência de propriedades globais que representam a cognição de fato. Isto posto, torna-se primordial compreender como as regras são formadas e consolidados no programa connexionista, a partir das técnicas de

⁴³ A base do sistema nervoso são os neurônios, que são células capazes de estabelecer conexões entre si ao receber estímulos do ambiente externo e interno do agente. Estes tipos de células são as responsáveis por transmitir impulsos nervosos ao cérebro e são classificadas em três tipos: sensoriais, motores e interneurônios. As sensoriais são responsáveis por levar a informação do ambiente externo – captadas por receptores mecânicos ou químicos – até o sistema nervoso. Os interneurônios – grupo mais numeroso – são aqueles responsáveis por transmitirem o sinal dos neurônios sensoriais ao sistema nervoso central. Por fim, os motores conduzem a resposta ao estímulo recebido, do sistema central ou órgão ou tecidos estimulados.

aprendizado não supervisionado, aprendizado supervisionado e aprendizagem de reforço.⁴⁴

Começando pela regra de Hebb, ou técnica de aprendizado não supervisionado⁴⁵, foi cunhada em 1949 pelo psicólogo Donald Olding Hebb e mostra que a formação das regras que condicionam a emergência de propriedades globais acontecem em duas fases distintas: aprendizado e aprendizado consolidado. A primeira fase acontece quando o sistema neural é exposto a uma nova situação, a qual motiva o conjunto de neurônios atuantes do processo a criarem/reforçarem suas conexões todas as vezes que essa operação – e com esta mesma configuração – aconteça. Nesta fase o número de neurônios atuantes é alto porque a nova regra ainda não está consolidada no sistema neural. Por sua vez, a segunda etapa de aprendizagem é caracterizada pela consolidação das regras pelos neurônios, ou seja, do estabelecimento da emergência das propriedades globais. Na etapa do aprendizado consolidado, o custo ao evocar a emergência global é infinitamente menor daquele necessário da fase de aprendizado.⁴⁶

Outra técnica de aprendizado usada pelas redes neurais é aquela conhecida como *backpropagation*, ou aprendizado supervisionado⁴⁷. Esta técnica consiste em treinar as unidades de entrada e demais unidades da rede para que as unidades de saída se moldem ao resultado desejado. Assim, através da recursividade aleatória e de ajustes nos pesos das unidades de entrada e demais unidades, o padrão de saída almejado se consolida no sistema.⁴⁸

⁴⁴ As três técnicas de aprendizado pertencem a uma área de estudos chamada *deep learning* – subárea do ramo Machine Learning – cujo objetivo último é treinar os computadores para realizar as mesmas tarefas executadas pelos seres humanos. Em geral, ao invés de regras pré-definidas, o *deep learning* visa que os sistemas de IA sejam capazes de aprender por eles próprios a partir do reconhecimento de padrões de informação.

⁴⁵ Traduzido do termo em inglês: *unsupervised*. Ou seja, a técnica natural de aprendizado encontrado na natureza, onde os agentes consolidam o aprendizado através da recursividade aleatória de padrões expostos à agência.

⁴⁶ Varela, Thompson, e Rosch, “The embodied mind”, 2016, 88.

⁴⁷ Traduzido do termo em inglês: *supervised*.

⁴⁸ « *Training nets to model aspects of human intelligence is a fine art. Success with backpropagation and other connectionist learning methods may depend on quite subtle adjustment of the algorithm and the training set. Training typically involves hundreds of thousands of rounds of weight adjustment. Given the limitations of computers presently available to connectionist researchers, training a net to perform an interesting task may take days or even weeks.* » James Garson, “Connectionism”, org. Edward N Zalta, Uri Nodelman, e Colin Allen, *The Stanford Encyclopedia of Philosophy*, 2018, 5.

O terceiro tipo de aprendizado que vem ganhando destaque nos últimos anos é aquele conhecido como aprendizagem de reforço⁴⁹, apesar de ainda permanecer uma grande diferença entre ele e as demais técnicas de *machine learning*⁵⁰. Esta abordagem consiste em operar de forma semelhante ao mimetismo praticado no reino animal, onde as recompensas e as punições envolvidas em um determinado evento ajudam a calibrar o aprendizado e a obter o resultado almejado. Esta técnica poderia explicar, por exemplo, os primeiros sucessos cognitivos alcançados por Hayy ao imitar os sons das gazelas para obter alimentos ou avisar sobre algum perigo iminente⁵¹.

As well as the different techniques in machine learning, there are three different types: supervised, unsupervised, and reinforcement learning. Supervised learning, which involves feeding a machine labeled data, is the most commonly used and also has the most practical applications by far. In the last few years, however, reinforcement learning, which mimics the process of training animals through punishments and rewards, has seen a rapid uptick of mentions in paper abstracts.⁵²

⁴⁹ Traduzido do termo em inglês: *reinforcement learning*.

⁵⁰ Ver Figura 2 – Total de menções em papers de reinforcement learning comparado a outras técnicas de machine learning, 33.

⁵¹ « Ele reproduzia também, com uma grande exatidão, todos os cantos de pássaros ou os gritos animais que ouvia. Mas reproduzia, sobretudo, os gritos das gazelas que pedem socorro ou que querem se aproximar uma da outra, ou que desejam alguma coisa, ou que procuram evitar um perigo: pois os animais, para essas ocasiões diferentes, têm gritos diferentes. » Ibn Tufayl, *O Filósofo Autodidata*, trad. Isabel Loureiro, Edição: 1 (São Paulo: UNESP, 2005), 62–63.

⁵² Karen Hao, “We Analyzed 16,625 Papers to figure out Where AI Is Headed Next”, *MIT Technology Review*, 2019, 9.

Reinforcement learning is gaining momentum

Share of papers that mention it compared to any type of machine learning

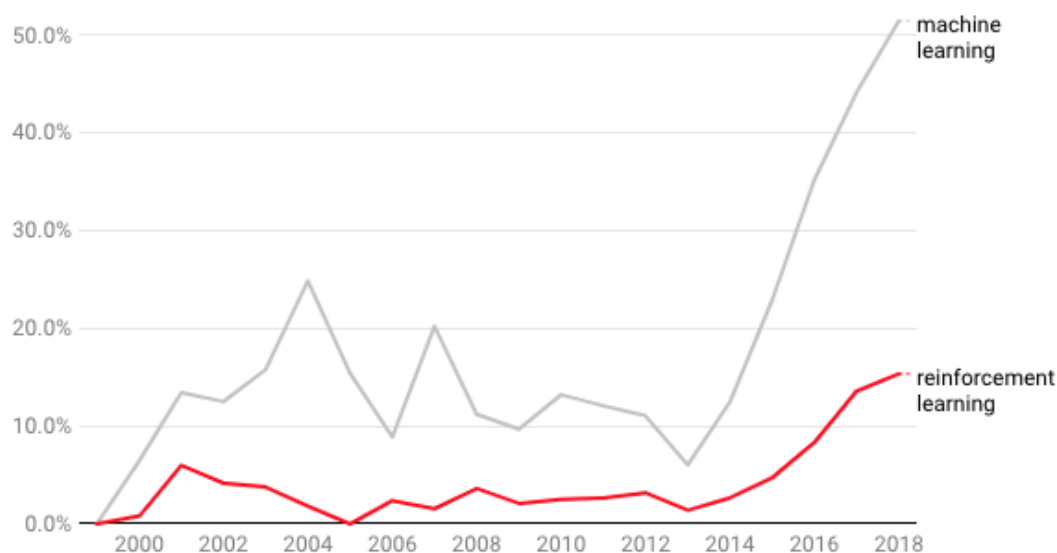


Chart: MIT Technology Review • Source: [arXiv.org](https://arxiv.org) • Created with Datawrapper

Figura 2 – Total de menções em papers de reinforcement learning comparado a outras técnicas de machine learning

Desse modo, a abordagem conexionista difere daquela proposta pelo cognitivismo por não impor ao sistema o uso de regras e símbolos para o incremento cognitivo. A capacidade plástica do cérebro e a propriedade de auto-organização dos neurônios permitem uma cognição dinâmica, flexível e independente, mostrando-se uma opção promissora para a autonomia cognitiva da agência.⁵³

⁵³ Apesar de a emergência das propriedades globais ser bem aceita entre os cientistas, ainda não há de facto uma teoria unificada. « *There is no unified formal theory of emergent properties. It is clear, however, that emergent properties have been found across all domains – vortices and lasers, chemical oscillations, genetic networks, developmental patterns, population genetics, immune networks, ecology, and geophysics.* » Varela, Thompson, e Rosch, “The embodied mind”, 2016, 88. Desse modo, a natureza da emergência das propriedades globais, assim como o problema relacionado a forma como a sintaxe acontece no cognitivismo, impõem-se como desafios e pontos sensíveis aos respectivos programas. Voltando as redes neurais, existem limitações que as teorias de aprendizado não respondem, como aquela conhecida como *one shot learning*, no qual os agentes consolidam o aprendizado em um único evento. Garson, “Connectionism”, 5.

Although classical systems are capable of multiple constraint satisfaction, connectionists argue that neural network models provide much more natural mechanisms for dealing with such problems.⁵⁴

Isto posto, ressalta que as técnicas de *deep learning* – aquelas baseadas em redes neurais –, têm-se mostrado mais eficazes nas tarefas da cognição do que aquelas mais clássicas, ou, mais especificamente, baseadas em regras lógicas em sistemas de IA. Conforme pode ser confirmado por uma pesquisa realizada pelo MIT, a frequência de palavras em *papers* como *theory* e *rule* aparecem em queda desde o ano 2000, e ao contrário, palavras como *learning* e *network* destacam-se com frequência alta, principalmente a partir do ano de 2007, corroborando a importância deste programa atualmente para o estudo da cognição em sistemas de IA.

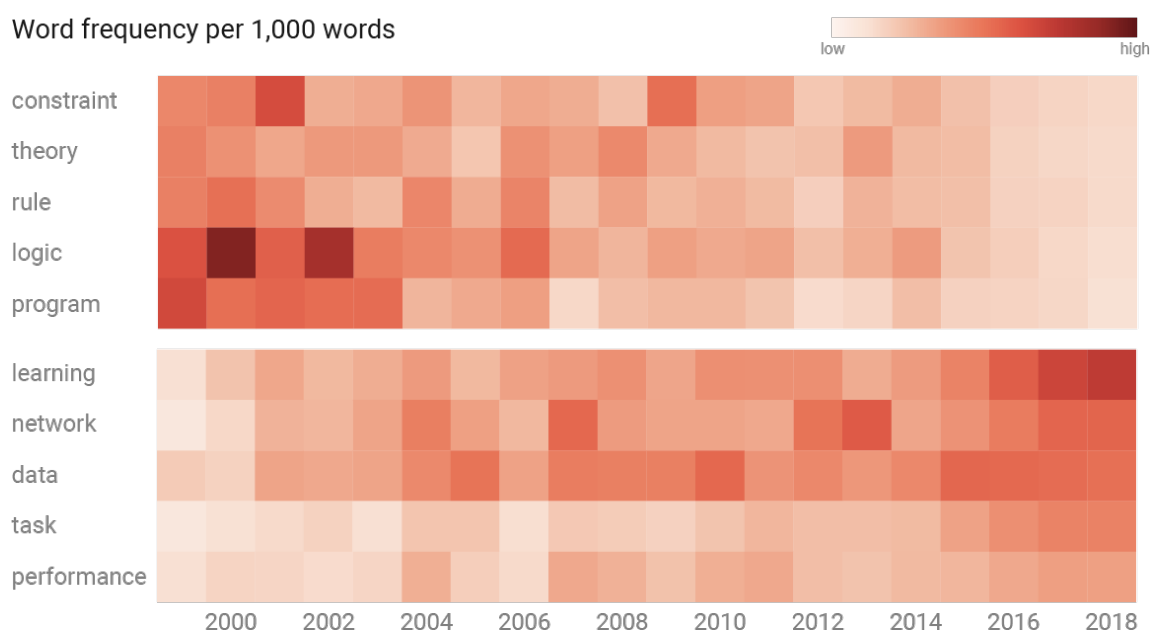


Chart: MIT Technology Review • Source: arXiv.org

Figura 3 – Gráfico de frequência de palavra a cada 1000 em papers entre os anos 2000 e 2018

Visto o que foi exposto sobre o connexionismo, pode-se afirmar que a cognição a partir deste programa é a emergência global de propriedades de uma rede composta pela miríade de simples componentes. Esta emergência se apresenta todas as vezes que é

⁵⁴ Garson, “Connectionism”, 9.

provocada por regras internas, não fornecidas pelo programador, mas desenvolvidas pelo histórico do agente. Assim, o sucesso da cognição é a apresentação da emergência global quando estimulada por um problema/constrangimento imposto à agência.

Consequentemente, a RM no programa conexionista assume características distintas daquelas propostas pelo programa cognitivista – na qual a RM é simbólica. Desse modo, a RM no conexionismo é manifestada por meio de a emergência de propriedades globais a partir de componentes neurais, sendo o valor semântico definido por funções distintas daqueles. Por causa dessas características este tipo de representação pode ser definido como RM subsimbólica.

Contudo, o programa conexionista não visa ser uma cisão completa com o programa cognitivista, pois é possível implementar alternativas e pontes entre ambos os programas, uma vez que o conexionismo não faz objeção aos atributos semânticos da RM, mas sim as características e a operação do símbolo no cognitivismo. Uma forma possível de complementaridade e integração entre ambos programas é exemplificado através da estrutura do DNA. As moléculas deste composto orgânico seriam compostos em um nível mais fundamental e básico de ácidos – que se comportariam como os componentes neurais – e as proteínas, em um nível descritivo superior corresponderiam a representação propriamente simbólica ou semântica.⁵⁵ Desse modo, esta abordagem apresenta estrutura subsimbólica no nível elementar e simbólica no nível descritivo – também conhecida como conexionismo implementacional.

[...] many connectionists do not view their work as a challenge to classicism and some overtly support the classical picture. So-called implementational connectionists seek an accommodation between the two paradigms. They hold that the brain's net implements a symbolic processor. True, the mind is a neural net; but it is also a symbolic processor at a higher and more abstract level of description. So the role for connectionist research according to the implementationalist is to discover how the machinery needed for symbolic processing can be forged from neural network materials, so that classical processing can be reduced to the neural network account.⁵⁶

Por fim, a RM do tipo [2], ou subsimbólica, será validada se todas as condições

⁵⁵ Garson, 9.

⁵⁶ Garson, 11.

determinadas nesta seção forem satisfeitas pela agência:

- I. O processamento mental é realizado de forma distribuída e não local;
- II. A atribuição semântica é realizada por funções distintas daquelas constituídas nos componentes neurais;
- III. A cognição acontece pela emergência global das propriedades da RM, quando provocada por regras locais previamente consolidadas pelo histórico da agência;
- IV. As técnicas de aprendizado que consolidam as regras neurais são: aprendizado não supervisionado (regra de Hebb), aprendizado supervisionado (*backpropagation*) e aprendizagem de reforço (*reinforcement learning*);
- V. A cognição é confirmada se a emergência global a partir das propriedades subsimbólicas permitirem à agência a resolução de algum problema/constrangimento imposto a ela.

1.3 Tipo 3 – a representação mental como ação incorporada (enativismo)

O terceiro tipo de programa de investigação que será apresentado é o enativismo. Este termo foi cunhado pelo neurobiólogo Humberto Maturana e o biólogo/filósofo Francisco Varela a partir da expressão espanhola *en acción*⁵⁷, definindo assim a essência deste programa: a emergência da cognição como ação perceptualmente guiada e consolidada pela recorrência do histórico de interações entre o agente – ou mais precisamente, do seu sistema sensoriomotor – e o mundo. Com relação a RM, pode-se apontar o enativismo como o mais radical dentre os três principais programas cognitivos, por romper com a necessidade da RM, pois cognição é ação incorporada pelo corpo – único canal de contato com o mundo. Desse modo, agente e mundo emergem juntos, mutuamente, num processo recíproco de especificação e consequentemente, refutando a necessidade de representação interna e externa no processo cognitivo.

Desta forma, o compromisso dos programas cognitivista e conexcionista com a RM, apresenta-se como uma objeção no programa enativista⁵⁸. Esta posição enativista retoma o clássico debate entre objetividade/subjetividade, ou em outras palavras, o velho e incontornável problema filosófico do ovo e da galinha. A posição da galinha afirma que há um mundo pré-estabelecido, onde a imagem é condição necessária para a cognição. Já a posição do ovo é aquela na qual o sistema cognitivo do agente projeta o seu próprio mundo, como um reflexo da constituição sistêmica da agência. Desse modo, o que veio antes, o mundo ou a imagem?⁵⁹ O enativismo apresenta uma resposta para

⁵⁷ Tradução para o português: em ação. O termo enativismo promove o diálogo entre tradições da ciência cognitiva, mais especificamente o conexionismo e a fenomenologia de Merleau-Ponty, e elementos da psicologia Budista – foco na ação presente e a desconstrução do ego empírico. O enativismo não deve ser confundido com a teoria do *embodied cognition*, que estuda a importância do corpo no processo cognitivo, porém sem assumir as mesmas diretrizes propostas pelo programa enativista.

⁵⁸ No programa conexcionista radical, não há a necessidade de símbolos no processo cognitivo, porém há uma correspondência entre a emergência do estado global de um determinado tipo e as propriedades do mundo que provocaram tal emergência. Já no cognitivismo há a representação explícita na mente de um agente de um objeto: o símbolo. Ao contrário desses programas, o programa enativista defende que a cognição não é representação de um mundo objetivo, mas sim a atuação do agente neste mundo. Atuação em vez de representação, portanto.

⁵⁹ Francisco J. Varela, Evan Thompson, e Eleanor Rosch, “Enaction: Embodied Cognition”, in *The Embodied Mind: Cognitive Science and Human Experience*, revised edition (Cambridge, Massachusetts ; London England: MIT Press, 2016), 172.

este impasse, um meio termo, intermediado pelo único canal possível entre o objetivo e o subjetivo: o corpo do agente.

Assim, a corporalidade da agência intermedia a troca de informações com o mundo, ou em outros termos, o mundo se apresenta através do apercebimento dos corpos dos agentes, e estes são especificados quando em contacto com este mundo. Dessa forma, o corpo é o meio do caminho entre as perspectivas realistas e idealistas da cognição.⁶⁰ A questão da corporalidade neste programa traz forte influência da filosofia de Merleau-Ponty, por exemplo, o caráter indissociável do corpo e do mundo, assim como a natureza textural sensível comum desta relação, tal como duas mãos que se tocam, alternando a configuração de tocar e ser tocada, imprimindo o que Merleau-Ponty chama «*double-sensations*», fenômeno que explicaria a falsa percepção de distinção do corpo dentre os demais objetos do mundo.⁶¹ Por consequência, se a percepção da distinção do corpo é falsa, a RM não se caracteriza como a encenação de algo independente do agente, mas sim da atuação – ou experiência – dele no mundo. Desse modo, a cognição se define não como a recuperação de propriedades extrínsecas ou intrínsecas, mas como a emergência de padrões incorporadas pela vivência da agência no mundo, sendo o processo emergencial orientado via percepção, resultando na ação de fato – daí a ideia de ação guiada perceptualmente.

Em vista disso, tem-se que o incremento cognitivo é dependente da estrutura corporal presente na agência. A percepção das cores, por exemplo, ocorre de forma particular entre os seres da natureza – quando se aponta para a cor «vermelha», a mesma

⁶⁰ No caso do Realismo, a cognição é a recuperação de um mundo pré-estabelecido, no qual a representação mental assume função de restabelecimento destas propriedades extrínsecas. No caso do idealismo, a cognição é a projeção do mundo interno do agente, no qual a representação mental assume a função de projetar internamente tal configuração. Varela, Thompson, e Rosch, 172.

⁶¹ « *When I press my two hands together, it is not a matter of two sensations felt together as one perceives two objects placed side by side, but of an ambiguous setup in which both hands can alternate the rôles of 'touching' and being 'touched'. What was meant by talking about 'double sensations' is that, in passing from one rôle to the other, I can identify the hand touched as the same one which will in a moment be touching. In other words, in this bundle of bones and muscles which my right hand presents to my left, I can anticipate for an instant the integument or incarnation of that other right hand, alive and mobile, which I thrust towards things in order to explore them. The body catches itself from the outside engaged in a cognitive process; it tries to touch itself while being touched, and initiates 'a kind of reflection' which is sufficient to distinguish it from objects.* » Maurice Merleau-Ponty, “The Experience of the Body and Classical Psychology”, in *Phenomenology of Perception* (London: Routledge, 2002), 106–7.

será percebida de formas diversas, a partir da configuração estrutural do sistema visual do perceptor. Assim, as células cones e bastonetes⁶², presentes na retina do sistema visual dos animais, são capazes de filtrar distintas intensidades do espectro luminoso, de acordo com a estrutura visual disponível em cada espécie animal: monocromática (animais marinhos e alguns mamíferos); dicromática (algumas espécies de primatas e boa parte dos mamíferos); tricromática (homem, algumas espécies de primatas e cangurus) e tetrocromática (répteis, anfíbios, aves e insetos). Portanto, a percepção da cor de cada indivíduo é resultado do seu tipo específico de estrutura visual, pelo que não seria possível – sob as lentes do enativismo – pensar numa cor vermelha objetiva/universal, pois ela é realmente percebida de múltiplas formas na natureza.⁶³

Não se trata apenas de qual sistema visual consegue captar mais cores, pois outros elementos sensoriais como por exemplo, o som e a percepção espacial, também influenciam na percepção das cores, ou seja, pode haver interdependências entre os sistemas perceptuais, fenômeno este conhecido como sinestesia. Como exemplo há o caso do artista Neil Harbisson que nasceu incapacitado de perceber cores desde o seu nascimento. Porém, devido a um dispositivo técnico acoplado em sua cabeça, Harbisson é capaz de reconhecer as cores através de frequências distintas de sons e produzir arte baseado neste modo de percepção⁶⁴. Há assim uma complexa interconexão entre estruturas de interação perceptuais que tornam a experiência de cada ser única no

⁶² As células cones – responsáveis pelo reconhecimento das cores – na espécie humana e alguns primatas é tricromática, ou seja, é capaz de reconhecer três frequências diferentes do espectro luminoso, correspondente às cores azul, vermelho e verde. Já os bastonetes são células responsáveis pela visão noturna, atuando na frequência correspondente a cor verde. “Cor e Visão – Carlos Fiolhais”, *Ano Internacional da Luz 2015* (blog), 22 de junho de 2015, <http://ail2015.org/index.php/2015/06/22/cor-e-visao-carlos-fiolhais/>.

⁶³ Ver Anexo 5 – Mecanismo tetrocromático versus mecanismo tricromático.

⁶⁴ Neil Harbisson, *I Listen to Color*, acessado 15 de abril de 2019, https://www.ted.com/talks/neil_harbisson_i_listen_to_color.

mundo.⁶⁵

Por isso, pode-se dizer que o mundo se apresenta por meio de diferentes percepções com os corpos, possibilitando-lhes moldarem o mundo e serem moldados por ele, imprimindo rastros destes eventos tanto em um quanto no outro. A cognição neste programa não é dependente somente da posição do corpo no mundo, mas também da forma com que cada sistema sensoriomotor o percebe. Assim, o incremento cognitivo envolve além das características biológicas, psicológicas e culturais⁶⁶, a atuação perceptual do sistema sensoriomotor, registrada na estrutura corporal e propagada na história filogenética dos seres vivos, permitindo-lhes dessa forma, o descortinar do mundo através da ação perceptualmente guiada, que é o que «*brings forth a world*».

As we can now appreciate, to situate cognition as embodied action within the context of evolution as natural drift provides a view of cognitive capacities as inextricably linked to histories that are lived, much like paths that exist only as they are laid down in walking. Consequently, cognition is no longer seen as problem solving on the basis of representations; instead, cognition in its most encompassing sense consists in the enactment or bringing forth of a world by a viable history of structural coupling.⁶⁷

Isto posto, a RM do tipo [3], ou enativista, será validada ao longo da tese somente se todas as condições determinadas nesta seção forem satisfeitas pela agência:

I. O sistema sensoriomotor, ou corpo, é condição necessária para o processo

⁶⁵ « *A natural proposal is that there are different experiential ways of having content corresponding to the different sense modalities. But it is far from given that the modalities mark out natural kinds of sensory experience, let alone the only natural kinds of such experience, since it is a complicated question how to distinguish the different kinds of sensory experience from one another. Furthermore, there are multi-sensory phenomena, such as synesthesia, which defy straightforward categorization as a single kind of sensory experience. Another problem case may be the experience of bodily actions, in which (say) visual, kinesthetic and tactile experiences are intertwined.* » Susanna Siegel, “The Contents of Perception”, in *The Stanford Encyclopedia of Philosophy*, org. Edward N. Zalta, Winter 2016 (Metaphysics Research Lab, Stanford University, 2016), <https://plato.stanford.edu/archives/win2016/entries/perception-contents/>.

⁶⁶ Além da dimensão biológica, a psicológica e a cultural também compõem o complexo processo cognitivo enativista. Como exemplo de influência da cultura nas cores, tem-se aquele relacionado a classificação léxica, o que consequentemente afeta a forma de interpretá-las. Na cultura inglesa há palavras distintas para as cores “blue” e “green”, contudo, a cultura Tarahumara – um povo nativo do México – usa o termo único para descrever o “verde” e o “azul”. Varela, Thompson, e Rosch, 171.

⁶⁷ Francisco J. Varela, Evan Thompson, e Eleanor Rosch, “Evolutionary Path Making and Natural Drift”, in *The Embodied Mind: Cognitive Science and Human Experience*, revised edition (Cambridge, Massachusetts ; London England: MIT Press, 2016), 205.

cognitivo.

- II. A cognição é a emergência guiada pelo sistema perceptual do agente que, de acordo com as respectivas características estruturais e históricas do sistema sensorimotor, descortina uma possibilidade, dentre várias possíveis, de aperceber e apresentar o mundo.
- III. O incremento cognitivo, ou consolidação dos padrões e emergências, dá-se via a atuação mútua entre o agente e o mundo. Sendo que tais interações ficam registradas tanto no corpo da agência quanto no mundo.
- IV. Não há RM no programa enativista, somente mútua especificação entre o agente e o mundo.
- V. A cognição é confirmada se o padrão emergencial apresentado pelos sistemas e subsistemas sensorimotores da agência, e consolidados devido ao seu histórico no mundo, permitirem à agência a resolução de algum problema/constrangimento imposto a ela.

Por fim, com os três tipos de RM definidas e suas condições de validação especificadas, torna-se possível enquadrar no próximo capítulo qual a RM necessária para a autonomia cognitiva em Hayy. Esta referência será usada como apoio no levantamento e averiguação das estruturas de RM do[s] tipo[s] [Y] necessárias para a autonomia cognitiva em sistemas de IA.

Capítulo 2 – O tipo de RM necessária para a autonomia cognitiva em Hayy

A obra de Ibn Tufayl, *O Filósofo Autodidata*, narra a jornada cognitiva e espiritual de Hayy Ibn Yaqzan, cujo nascimento tem duas versões obscuras na trama: a primeira de que teria sido gerado de forma espontânea por condições naturais e a segunda versão, em que Hayy teria sido fruto de um amor proibido. Independente da causa de sua origem, o fato é que o destino de Hayy é inevitavelmente o mesmo: o aflorar em uma ilha deserta e florescer cognitivamente.

Assim, esta obra clássica da filosofia árabe do século XII, desperta a curiosidade dos leitores não somente por ser um livro com nove séculos de transcurso histórico, mas também pelo contexto particular no qual o protagonista é colocado – e que leva ao despertar da consciência intelectual e espiritual, advinda do processo contemplativo e investigativo de um ser que parte da completa ignorância para a sabedoria e plenitude espiritual.

Pode-se interpretar a saga de Hayy Ibn Yaqzan como um experimento mental, composto de três dimensões distintas: a espiritual, a social e a cognitiva. Esta última dimensão será aquela que fornecerá a lente para o exame da relação entre tipos de RM e autonomia cognitiva na agência. Sob o prisma da obra, mais especificamente da sua dimensão cognitiva, o que se pode inferir da jornada de Hayy? Como é possível que uma criança sozinha, sem saber falar⁶⁸ e sem acesso a nenhum tipo de informação, conseguisse a proeza de se tornar um sábio? Qual o tipo de RM utilizada por Hayy para chegar a gozar de tal autonomia cognitiva? Será tal tipo de RM aquele necessário para a cognição em sistemas de IA?

Isto posto, e com objetivo de averiguar tais questões, esta tese irá focar na

⁶⁸ A linguagem não é algo que possa ser desprezado no processo cognitivo. Um estado mental geralmente é constituído pelo relacionamento entre RM e atitudes proposicionais. Esta última será aquela que concederá valor semântico a RM e levará o agente a agir. Como exemplo, pode-se imaginar o objeto coco sob a forma de uma RM e uma atitude proposicional como Hayy acreditar que cocos são apetitosos. Contudo, na trama de Tufayl, Hayy somente irá aprender a falar na sétima fase da vida – 49 anos. Esta tese não visa analisar a obra *O Filósofo Autodidata*, mas vale a pena apontar que a questão da fala poderia se apresentar como um problema, ou no mínimo como um fator duvidoso na trajetória cognitiva de Hayy.

primeira etapa da jornada cognitiva de Hayy – das sete especificadas no livro. Neste estágio específico, que abrange o período do surgimento do protagonista na ilha até a idade de 7 anos, Hayy vive a primeira infância, o florescimento do conhecimento sensível, a simpatia e a afeição dos seres que se lhe assemelham.

2.1 A primeira fase cognitiva de Hayy

Os trechos escolhidos para averiguar a relação entre a cognição e a RM em Hayy, descrevem em ordem sequencial, primeiro o florescimento dos instintos de sobrevivência do protagonista, aprendidos no convívio com as gazelas⁶⁹ – dentre as quais, a sua mãe adotiva – e com os demais animais da ilha. Pode-se pensar no caráter mimético – reprodução maquinal de gestos e atitudes – desta etapa cognitiva.

Ele reproduzia também, com uma grande exatidão, todos os cantos de pássaros ou os gritos de outros animais que ouvia. Mas reproduzia, sobretudo, os gritos das gazelas que pedem socorro ou que querem se aproximar uma da outra, ou que desejam alguma coisa, ou que procuram evitar um perigo: pois os animais, para essas ocasiões têm gritos diferentes. Os animais e ele se conheciam e não se tratavam como estranhos.⁷⁰

Em seguida, apresentam-se as primeiras RM em Hayy.

Com o tempo, passou a se lembrar das coisas mesmo quando elas estavam ausentes; percebeu que umas lhe inspiravam desejo, outras aversão.⁷¹

Aqui surge o despertar do protagonista para as particularidades do seu corpo comparado com os outros animais da ilha. No transcorrer dos anos, Hayy fica cada vez mais frustrado por não possuir os mesmos atributos físicos dos outros bichos: chifres, pelos, garras, etc., colocando-o sempre em desvantagem competitiva na disputa de alimentos. Esta etapa marca também a recorrência reflexiva das experiências do agente.

Enquanto isso, observava todos os animais [...]. Observava sua rapidez na corrida, sua força, as armas de que se haviam munido para lutar contra o adversário, tais como

⁶⁹ « O menino não parou, assim, de viver com as gazelas, reproduzindo os seus gritos com a voz, a ponto de ser confundido com elas. » Ibn Tufayl, *O Filósofo Autodidata*, trad. Isabel Loureiro, Edição: 1 (São Paulo: UNESP, 2005), 62.

⁷⁰ Tufayl, 63.

⁷¹ Tufayl, 63.

chifres, presas, cascos, esporões, garras. Voltando a si mesmo, via-se nu e sem armas, lento na corrida, fraco contra os animais que lhe disputavam os frutos, se apropriavam deles em seu detrimento e lhos tiravam sem que pudesse enxotá-los ou escapar-lhes. [...] Em si mesmo não constatava nada disso e, por mais que refletisse, não descobria a causa.⁷²

Finalmente, depois de experienciar e refletir sobre as aflições que lhe incomodavam, Hayy decide fabricar um bastão com galhos de árvores e produz então o seu segundo artefato – o primeiro foi uma espécie de roupa feita de folhas de palmeiras⁷³ – consolidando, neste momento, as suas capacidades cognitivas.

Todas essas constatações lhe eram penosas e o afligiam. A tristeza que sentiu por causa disto durou muito tempo [...]. Fabricou bastões com galhos de árvores, que poliu nas extremidades e alisou uma ponta à outra, e os brandia contra os animais com que tinha de lutar, atacando os fracos e resistindo aos fortes.⁷⁴

2.2 O experimento do bastão

Depois da exposição dos excertos das primeiras experiências cognitivas de Hayy, esta seção terá a incumbência de esquematiza-los afim de que se possa propor um experimento mental capaz de validar tanto o tipo de RM utilizada por Hayy quanto aquelas que serão averiguadas no Capítulo 3 – As RM do tipo [Y] necessárias para a autonomia cognitiva em sistemas de IA.

A partir do acontecimento da produção do bastão defensivo, pode-se organizar este fluxo em quatro etapas (ET) consecutivas.

ET1 – Motivo:

(i) os animais que lhe disputavam os frutos, se apropriavam deles em seu detrimento e lhos tiravam sem que pudesse enxotá-los ou escapar-lhes.⁷⁵

ET2 – Posse de RM com conteúdo semântico:

⁷² Tufayl, 63–64.

⁷³ « E já estava perto dos sete anos quando, sem esperança de ver realizar-se nele as vantagens naturais cuja ausência o fazia sofrer, tomou de grandes folhas de árvores, pondo-as umas atrás, outras à frente, e as prendeu a uma espécie de cinto que fez em volta da cintura com folhas de palmeira e de esparto. » Ibn Tufayl, *O Filósofo Autodidata*, trad. Isabel Loureiro, Edição: 1 (São Paulo: UNESP, 2005), 64.

⁷⁴ Tufayl, 64–65.

⁷⁵ Tufayl, 63.

- (ii) Com o tempo, passou a se lembrar das coisas mesmo quando elas estavam ausentes;
 (iii) percebeu que umas lhe inspiravam desejo, outras aversão.⁷⁶

ET3 – Processamento recorrente das instâncias de RM:

- (iv) por mais que refletisse, não descobria a causa.⁷⁷ [...] (v) A tristeza que sentiu por causa disto durou muito tempo [...].⁷⁸

ET4 – Incremento cognitivo:

- (vi) Fabricou bastões com galhos de árvores, que poliu nas extremidades e alisou uma ponta à outra, (vi) e os brandia contra os animais com que tinha de lutar, atacando os fracos e resistindo aos fortes.

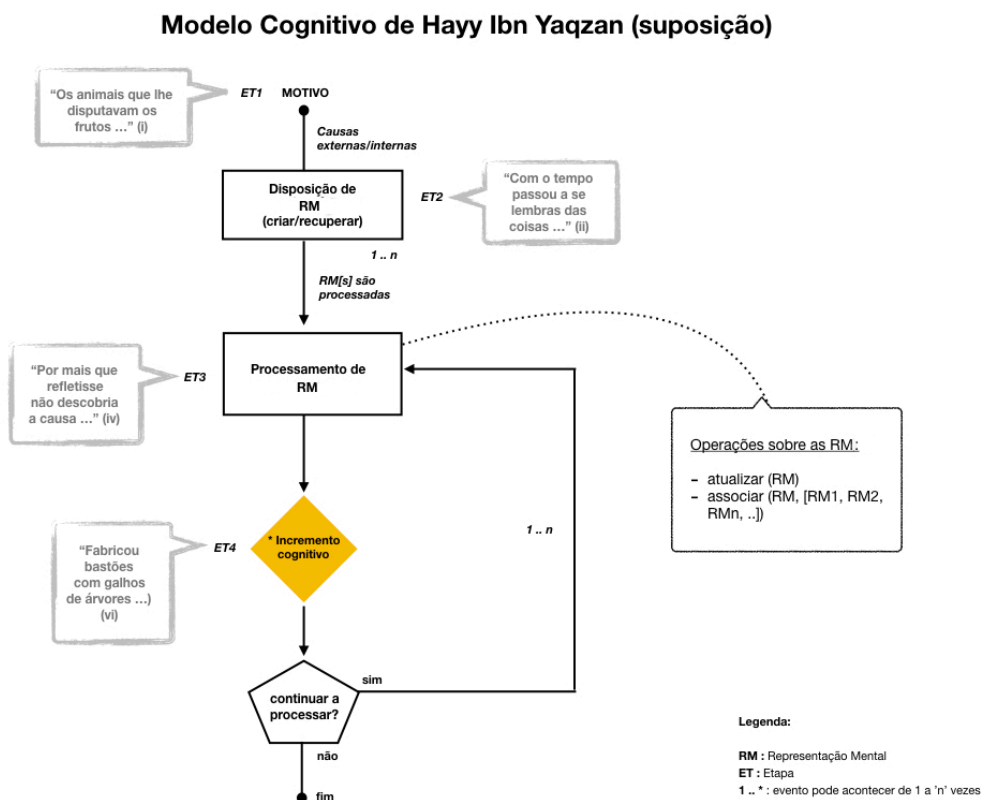


Figura 4 – Suposto modelo cognitivo de Hayy, baseado na primeira fase de sua vida – 0 aos 7 anos.

O objetivo do experimento do bastão é averiguar se sistemas de IA seriam capazes

⁷⁶ Tufayl, 63.

⁷⁷ Tufayl, 63–64.

⁷⁸ Tufayl, 64–65.

de atingir IV sem a necessidade de passar por ET2 e ET3, ou em outras palavras, dada um motivo para o sistema de IA, I, que ele possa construir um bastão, IV, sem operacionalizar RM – ET2 e ET3.

O sistema de IA deverá ainda ser capaz de realizar este incremento de forma autônoma, ou seja, independente de qualquer intervenção técnica que pudesse auxiliá-lo na tarefa e, portanto, submetendo-o as mesmas restrições encontradas por Hayy na trama de Tufayl – completo isolamento físico e sem acesso a orientação e atualização – informacional ou estrutural – de qualquer tipo. O sucesso do experimento será alcançado quando os agentes de IA forem capazes de atingir o estado final, cumprindo todas as condicionantes supracitadas.

As próximas secções deste capítulo irão averiguar quais dos tipos de RM especificados no Capítulo 1 – A representação mental explicam o êxito cognitivo de Hayy no experimento do bastão defensivo.

2.3 O experimento do bastão sob as lentes do cognitivismo

Na secção 1.1 Tipo 1 – a representação mental simbólica (cognitivismo) foi possível averiguar que o tipo [1] – ou tipo simbólico – explica o funcionamento da mente através das seguintes condições:

- I. O funcionamento da mente assemelha-se com uma máquina de Turing (MT);
- II. Os agentes devem ser capazes de possuir RM;
- III. A cognição é operacionalizada sintaticamente, através da computação das propriedades físicas e não das propriedades semânticas de símbolos;
- IV. A cognição é confirmada se após essa computação, a agência tiver êxito no problema/constrangimento imposto a ela.

Consequentemente, resta saber se as condições deste programa satisfazem aquelas definidas para o experimento do bastão.

Com relação a condição I, a mente de Hayy seria semelhante a uma MT. Logo, pode-se supor que dentre as várias MT possuídas pelo agente, uma fosse específica ao

conteúdo do desejo de comer frutas.⁷⁹ De forma simplória, poder-se-ia esboçar os estados (S) desta MT com a seguinte configuração:

- S1. desejo comer uma fruta
- S2. disputar alimento com outros animais
- S3. não desejo comer fruta

Segundo os itens (i) e (v) do experimento do bastão, Hayy conviveu muito tempo com a frustração de não vencer a disputa por comida com os outros animais, o que consequentemente criou uma recorrência entre os estados S1 e S2 da máquina. Desse modo, Hayy quase nunca atingia o estado S3, estado que representaria a satisfação do desejo e a atualização semântica para o símbolo “fruta”. O surgimento desta máquina poderia ser explicado devido ao processo mimético adquirido no convívio com as gazelas, já que tratam de estados básicos de sobrevivência comuns a todos os animais, e presente no processo cognitivo de várias espécies: primatas, felinos, pássaros, etc.⁸⁰

Contudo, no contexto do experimento do bastão, Hayy modifica a função S2 por uma nova, que será chamada de S2': disputar alimentos com outros animais usando bastão. Vale lembrar que numa máquina de estados, o que determina a mudança de estado são as funções de transição, as quais compõem a tabela de máquina. Desse modo, pode-se supor que os estados presentes antes do incremento da cognição, teriam sido aprendidos via mimetismo, como por exemplo, morder o oponente. Porém, a partir da substituição de S2 por S2', ele alterou a MT para além do que o instinto ou o mimetismo lhe permitiam.

Todavia, isto traz uma questão não tratada pelo programa cognitivista: a de que um agente seria capaz de elaborar as suas próprias MT. Em sistemas de IA este procedimento é concretizado por desenvolvedores – porém, o que determina isto na

⁷⁹ Uma máquina de Turing pode ter de 1 a 'n' fitas – espaço de armazenamento –, sendo que todas as fitas poderiam ser reorganizadas em uma única fita. De Mol Liesbeth, “Turing Machines”, org. Zalta, *The Stanford Encyclopedia of Philosophy*, nº Winter 2018 (2018): 32, <https://plato.stanford.edu/archives/win2018/entries/turing-machine/>. Portanto, seria possível que um agente portasse múltiplas MT, cada uma correspondente a um tipo de conteúdo, podendo estarem associadas a 1 ou 'n' RM simbólicas.

⁸⁰ A questão de MT inerentes e/ou auto implementáveis nos agentes biológicos apresenta novas possibilidades de pesquisa e poderiam ser uma forma de propor uma explicação ao problema relacionado a sintaxe/semântica no cognitivismo. Porém, estes pontos estão além do escopo definido por esta tese.

agência biológica? Este é justamente um dos grandes desafios que devem ser explicados pelo cognitivismo.

As we mentioned in a computer program the syntax of the symbolic code mirrors or encodes its semantics. In the case of human language, it is far from obvious that all of the semantic distinctions relevant in an explanation of behavior can be mirrored syntactically. [...] Furthermore, although we know where the semantic level of a computer's computations comes from (the programmers), we have no idea how the symbolic expressions supposed by the cognitivist to be encoded in the brain would get their meaning.⁸¹

Logo, a única forma de solucionar este impasse referente ao experimento seria:

- I – Supor que os agentes são capazes de auto implementarem MT;
- II – Admitir a interferência/atualização externa da função de transição.

O item I não é suportado pela agenda cognitivista e o item II viola a condicionante de autonomia exigida no experimento. Por estas razões, o programa cognitivista não explica o sucesso cognitivo de Hayy no experimento do bastão.

2.4 O experimento do bastão sob as lentes do connexionismo

Na seção 1.2 Tipo 2 – a representação mental subsimbólica (conexionismo) foi possível averiguar que o tipo [2] de RM, ou subsimbólico, explica o funcionamento da mente através das seguintes condições:

- I. O processamento mental é realizado de forma distribuída e não local;
- II. A atribuição semântica é realizada por funções distintas daquelas constituídas nos componentes neurais;
- III. A cognição acontece pela emergência global das propriedades da RM, quando provocada por regras locais previamente consolidadas pelo histórico da agência;
- IV. As técnicas de aprendizado que consolidam as regras neurais são: aprendizado não supervisionado (regra de Hebb), aprendizado supervisionado (*backpropagation*) e aprendizagem de reforço

⁸¹ Varela, Thompson, e Rosch, “The embodied mind”, 2016, 42.

(*reinforcement learning*);

- V. A cognição é confirmada se a emergência global a partir das propriedades subsimbólicas permitirem à agência a resolução de algum problema/constrangimento imposto a ela.

Resta assim saber se as condições deste programa satisfazem aquelas definidas para o experimento do bastão.

Com relação a condicionante I, o processamento distribuído das RM não apresenta um empecilho para as condições impostas no experimento do bastão, uma vez que a vertente connexionista implementacional acomoda também as RM simbólica e, portanto, concorda com as condicionantes do experimento ET2 – Posse de RM com conteúdo semântico – e ET3 – Processamento recorrente das instâncias de RM. Contrariamente, caso a escolha fosse a vertente connexionista radical, ou seja, aquela que nega a RM simbólica no processo mental, não seria possível contemplar as condicionantes ET2 e ET3, e consequentemente, este programa seria incapaz de explicar a autonomia cognitiva no caso de Hayy.⁸²

Em relação a condicionante II, o facto de as funções semânticas não serem realizadas pelos componentes neurais, também não impede o fluxo no experimento do bastão, uma vez que há o compromisso da vertente connexionista implementacional de acomodar as RM simbólicas, conforme explicado no paragrafo anterior. Contudo, ressalta-se que ambas vertentes connexionistas – radical e implementacional – precisam explicar a operacionalidade do processo de atribuição semântica das RM subsimbólicas, com ou sem o recurso às RM simbólicas.

A condicionante III se acomoda de forma satisfatória no contexto do experimento. Afinal, as condicionantes ET1, ET2 e ET3 fundamentam a história cognitiva de Hayy no episódio da disputa das frutas. Deste modo, o acesso às RM via processos não

⁸² O connexionismo radical justificaria o experimento do bastão por meio somente das condicionantes ET1 e ET4. De forma breve, pode-se explicar esta dinâmica da seguinte forma: a recorrência da ET1 durante um tempo indeterminado consolidaria um novo padrão global e respectivamente, atingir-se-ia a ET4. Contudo, isto vai contra o exposto pelo enredo da obra que explicita o uso da RM no evento do bastão. Portanto, a escolha do connexionismo implementacional está relacionado às circunstâncias da trama *O Filósofo Autodidata* e também a forte evidência do uso da RM na cognição humana. Por fim, é válido ressaltar que neste capítulo as ET2 e ET3 são condicionantes necessárias no enquadramento do tipo de RM em Hayy, ao contrário do próximo capítulo, no qual será desejável que os sistemas de IA aufram a ET4 sem percorrer por ET2 e ET3.

sensoriais e não presenciais, ET2, operacionalizadas através de recorrências não presenciais, ET3, explicam a dinâmica de consolidação da história em Hayy, assim como o assentamento do padrão global incorporado pela mente do agente, apresentando-se como emergência global sempre que o mesmo padrão consolidado for provocado por fatores externos ou internos.

A condicionante IV, por sua vez, propõe métodos de aprendizado que permitem a um agente na condição de Hayy, ou seja, completa ignorância, a evolução paulatina e autônoma da cognição. Logo, todas as técnicas de aprendizado neurais poderiam ser usadas para explicar a autonomia cognitiva em Hayy, principalmente a técnica de aprendizado supervisionado, que sob as lentes do conexionismo implementacional fornece uma teoria robusta para explicar o experimento do bastão.

Assim, o modelo de aprendizado *backpropagation*, ou método supervisionado de aprendizado, é consonante com a etapa ET3 – Processamento recorrente das instâncias de RM –, no experimento do bastão. Dado que Hayy é capaz de produzir RM⁸³ e ajustá-las mentalmente ao resultado esperado⁸⁴, pode-se supor que a cada recorrência mental – recorrência aleatória –, Hayy experimenta diversas formas de vencer a disputa pelas frutas com os outros animais – treino das unidades de entrada –, sendo que uma destas recorrências fosse aos poucos se consolidando de forma mais eficaz do que as outras – ajustes dos pesos dos dados de entrada neurais –, delineando o padrão mental que resultará no assentamento da emergência global e no incremento cognitivo: a produção do bastão, ou em outras palavras, a saída almejada do aprendizado supervisionado. Nesta situação, a dinâmica entre o método conexionista implementacional e a técnica de aprendizado supervisionado, parecem fornecer uma teoria poderosa para explicar a autonomia cognitiva em Hayy.

⁸³ ET2 – Posse de RM com conteúdo semântico.

⁸⁴ Esta tese propõe que na etapa do experimento ET3, – Processamento recorrente das instâncias de RM, aconteça o processamento da cognição de fato. Assim, a dinâmica do programa conexionista implementacional, suportaria a posse de RM simbólicas – descrição *higher level* – cujo processamento distribuído estaria a cargo dos componentes neurais – RM subsimbólicas. Desse modo, estas seriam as responsáveis pelo processamento da informação que será atribuída as RM simbólicas. Caberia investigar a natureza e o funcionamento dos componentes neurais na mente humana e nos outros animais, como os primatas, por exemplo, e encontrar nos respectivos contrastes elementos que expliquem o porquê da dissimilaridade desta funcionalidade entre as espécies.

Por fim, estando em acordo as condicionantes da RM do tipo [2], ou subsimbólica, com as condicionantes definidas no experimento do bastão, esta tese assume a posição de que a RM do tipo conexionista implementacional e a técnica de aprendizado supervisionado⁸⁵ explicam e atendem as demandas exigidas pelo experimento do bastão e, consequentemente, a autonomia cognitiva em Hayy.

2.5 O experimento do bastão sob as lentes do enativismo

Na seção 1.3 Tipo 3 – a representação mental como ação incorporada (enativismo) foi possível averiguar que o tipo [3] de RM, ou enativista, explica o funcionamento da mente de acordo com as seguintes condições:

- I - O sistema sensorimotor, ou corpo, é condição necessária para o processo cognitivo;
- II - A cognição é a emergência guiada pelo sistema perceptual do agente, que de acordo com as respectivas características estruturais e históricas do sistema sensorimotor, descortina uma possibilidade, dentre várias possíveis, de aperceber e apresentar o mundo;
- III - O incremento cognitivo ou a consolidação dos padrões emergentes, dá-se via a atuação mútua entre agente e mundo. Sendo que tais interações ficam registradas tanto no corpo do agente quanto no mundo;
- IV - Não há RM no programa enativista, somente mútua especificação entre o agente e o mundo;
- V - A cognição é confirmada se o padrão emergencial apresentado pelos sistemas e subsistemas sensorimotores do agente, e consolidados devido ao seu histórico no mundo, permitirem à agência a resolução de algum problema/constrangimento imposto a ela.

Neste programa específico, há uma condicionante que antecipadamente impossibilita o ajuste deste programa ao experimento do bastão: a condicionante IV,

⁸⁵ Corroborando com este argumento, cita-se a pesquisa do MIT apontando a técnica do aprendizado supervisionado como a mais usada dentre todas as técnicas de Machine Learning. «Supervised learning, which involves feeding a machine labeled data, is the most commonly used and also has the most practical applications by far.» Karen Hao, “We Analyzed 16,625 Papers to figure out Where AI Is Headed Next”, 2019, 8.

uma vez que não há RM no programa enativista. De acordo com as condicionantes ET2 – Posse de RM com conteúdo semântico – e ET3 – Processamento recorrente das instâncias de RM –, é necessário que o agente seja capaz de possuir RM, o que não é aceito no programa enativista. Isto posto, não seria possível explicar a cognição em Hayy exclusivamente pelas interações do seu corpo com o ambiente – a ilha.

internal representations are used to instruct motor programs, which are essentially separate and independent from cognition. Hence, cognitive processing is not inextricably shaped by bodily actions.⁸⁶

Outra condicionante que sofreria uma forte objeção é a III – O incremento cognitivo ou consolidação dos padrões emergentes, dá-se via a atuação mútua entre o agente e o mundo. O excerto do experimento apresenta fortes indícios de que a cognição não aconteceu somente na interação com o mundo, pois as RM tiveram forte papel na cognição em Hayy, conforme os excertos do experimento em ET2:

(ii) «Com o tempo, passou a se lembrar das coisas mesmo quando elas estavam ausentes»⁸⁷;

(iii) «percebeu que umas lhe inspiravam desejo, outras aversão»⁸⁸;

e em ET3:

(iv) «por mais que refletisse, não descobria a causa.»⁸⁹;

(v) «A tristeza que sentiu por causa disto durou muito tempo [...]».⁹⁰

Pode ainda citar-se a importância da fofoca e da criação dos mitos no processo cognitivo humano, tal como descrito no livro *Sapiens*.⁹¹ A capacidade dos seres humanos criarem e transmitirem informações sobre coisas que não aconteceram de fato seria uma das razões pelas quais milhões de indivíduos atuam juntos em prol do mesmo objetivo. Deste modo, a cooperação entre complexas e numerosas sociedades seria

⁸⁶ Robert A. Wilson e Lucia Foglia, “Embodied Cognition”, org. Edward N Zalta, Uri Nodelman, e Colin Allen, *The Stanford Encyclopedia of Philosophy*, Spring Edition de 2017, 24.

⁸⁷ Tufayl, *O Filósofo Autodidata*, 63.

⁸⁸ Tufayl, 63.

⁸⁹ Tufayl, 63–64.

⁹⁰ Tufayl, 64–65.

⁹¹ Yuval Noah Harari, “The Tree of Knowledge”, in *Sapiens: a brief history of humankind* (United States, 2015), 36.

sustentada em sua grande maioria por objetivos abstratos, mitos e narrativas sobre coisas que nunca foram percebidas no mundo.⁹² Por fim, o programa enativista, apesar de apontar a relevante questão das características corporais no processo cognitivo da agência, parece ser o mais frágil quanto aos três tipos de RM para explicar a autonomia cognitiva em Hayy .

Yet the truly unique feature of our language is not its ability to transmit information about men and lions. Rather, it's the ability to transmit information about things that do not exist at all. As far as we know, only Sapiens can talk about entire kinds of entities that they have never seen, touched or smelled.⁹³

⁹² « Legends, myths, gods and religions appeared for the first time with the Cognitive Revolution. Many animals and human species could previously say, 'Careful! A lion!' Thanks to the Cognitive Revolution, Homo sapiens acquired the ability to say, 'The lion is the guardian spirit of our tribe.' This ability to speak about fictions is the most unique feature of Sapiens language. » Yuval Noah Harari, "The Tree of Knowledge", in Sapiens: a brief history of humankind (United States, 2015), 42.

⁹³ Harari, 41–42.

Capítulo 3 – As RM do tipo [Y] necessárias para a autonomia cognitiva em sistemas de IA

Neste capítulo serão apresentados autores antirrepresentacionistas que defendem programas de investigação acerca da cognição sem recurso a RM. Assim, a partir de textos técnicos e filosóficos serão expostas abordagens teóricas e principalmente práticas, respaldadas por projetos que implementaram robôs e sistemas de IA.

A primeira exposição ficará a cargo do roboticista Rodney Brooks, que irá apresentar um modelo de inteligência que dispensa o uso da representação informacional. Os robôs de Brooks conseguem interagir com o ambiente de forma independente, aparentemente sem a necessidade de regras que envolvem formas de representação. Contudo, tal proposta também conseguirá atender as demandas da autonomia cognitiva?

3.1 As criaturas de Rodney Brooks – inteligência sem representação

Rodney Brooks é roboticista e ocupa o cargo de diretor do laboratório de inteligência artificial e ciência da computação do MIT. Contudo, Brooks não atuou somente na esfera acadêmica, ele também foi diretor chefe técnico da empresa IRobot⁹⁴ e fundou a empresa pioneira de robótica *Rethink Robotics*⁹⁵.

Brooks no texto *Intelligence without representation*⁹⁶ apresenta uma proposta de inteligência – encarada como habilidade para entender e solucionar problemas – sem o artifício das RM. No entanto, como isto é possível? Como um robô consegue interagir em um ambiente dinâmico sem o uso de nenhum tipo de representação de informação, banco de dados ou mesmo regras condicionais?

Brooks começa a explicar a sua teoria através de uma digressão biológica e evolutiva sobre a inteligência. No decorrer da história do planeta, a inteligência é uma

⁹⁴ “IRobot Vacuum Cleaning, Mopping & Outdoor Maintenance”, acessado 11 de fevereiro de 2019, <https://www.irobot.com/>.

⁹⁵ A empresa *Rethink Robotics* foi adquirida em 2018 pelo grupo alemão HAHN. “Rethink Robotics”, in *Wikipedia*, 5 de janeiro de 2019, <http://bit.ly/2VyXzJw>.

⁹⁶ Brooks, “Intelligence Without Representation”.

capacidade recente comparada com outras habilidades biológicas, conforme pode-se constatar pelos seguintes marcos biológicos: 4.6 bilhões de anos (idade aproximada do planeta Terra); 3.5 bilhões de anos (entidades de células simples); 2.5 bilhões de anos (plantas fotossintéticas); 550 milhões de anos (vertebrados e peixes); 450 milhões de anos (insetos); 370 milhões de anos (répteis); 330 milhões de anos (dinossauros); 250 milhões de anos (mamíferos); 120 milhões de anos (primeiros primatas); 18 milhões de anos (macacos); 2.5 milhões de anos (homem); 10.000 anos (agricultura); 5.000 anos (escrita).⁹⁷

Desta forma, a natureza passou boa parte de sua existência desenvolvendo a mobilidade, a visão e a habilidade de sobrevivência dos seres em ambientes complexos e dinâmicos. A inteligência foi uma competência que surgiu muito mais tarde e, segundo o autor, seria menos complexa comparada com as habilidades de mobilidade e interação com o ambiente. Consequentemente, o que se aprende com a história da evolução biológica é que a pesquisa dos autômatos deve focar primeiro no desenvolvimento das capacidades básicas de interação, mobilidade e percepção do ambiente, para então numa etapa posterior almejar a complexa empreitada de replicar a mente humana.⁹⁸

Em Brooks, a abstração é o processo que gera os desdobramentos teóricos/conceituais das descrições do mundo, ou seja, a parte difícil e complexa da inteligência⁹⁹. Assim, um dos problemas relacionado com as abstrações é a sua relação com a leitura subjetiva daquilo que se considera um conceito, e outro problema é o do compromisso com a representação objetiva do mundo. Isto traz enorme complexidade de implementação para o universo dos robôs. Os cientistas criam abstrações do mundo e os implantam em sistemas de IA supondo que tais conteúdos sejam compostos de

⁹⁷ Brooks, 140.

⁹⁸ « *we will never understand how to decompose human level intelligence until we've had a lot of practice with simpler level intelligences.* » Rodney A. Brooks, "Intelligence Without Representation", *Artificial Intelligence* 47, nº 1-3 (janeiro de 1991): 139, [https://doi.org/10.1016/0004-3702\(91\)90053-M](https://doi.org/10.1016/0004-3702(91)90053-M).

⁹⁹ É justamente a capacidade de abstração, aquela que é muito superior nos seres humanos comparada com outros animais. Logo, parece que esta capacidade está atrelada a organização da agência humana, matéria que está sendo averiguada nesta tese.

blocos descritivos, «*block worlds*»¹⁰⁰, confiáveis e bem entendidos, aptos a serem capturados no mundo e remontados nos modelos abstratos que irão levar ao desempenho dos robôs.

Por isso, o autor considera esta abordagem científica ineficaz e ineficiente para o progresso da IA, pois a abstração em sistemas de IA é de fato baseada nos modelos concebidos pela mente humana. Dentro desta perspectiva, os sistemas de IA nada mais fazem do que computar entradas e saídas de dados produzidos a partir de modelos concebidos pelo homem. Portanto, não se trata de uma inteligência autônoma, capaz de realizar a percepção do ambiente e resolver problemas por si própria, mas sim de um ser que encena as representações definidas pelos cientistas e programadores, ou em outras palavras, uma redução da mente humana. Diante disto, poder-se-ia perguntar se os desempenhos de sistemas de IA não seriam de fato apenas os desempenhos dos próprios homens.

Assim, a condição para um robô possuir inteligência autônoma é poder atuar sem o artifício de RM – ou abstrações – do mundo. A inteligência se apresenta, antes de tudo, como a capacidade de percepção e interação com o ambiente. Eis então a especificação daquilo que Brooks¹⁰¹ nomeará de criaturas.

A abordagem assemelha-se à abordagem do tipo de RM apresentada na seção 1.3 Tipo 3 – a representação mental como ação incorporada (enativismo), podendo constituir uma objeção à necessidade da RM para a autonomia cognitiva da agência. Retornando a proposta de Brooks, resta saber como a arquitetura técnica de suas criaturas explicam a autonomia da inteligência e averiguar se tal modelo atende as particularidades da cognição.

Isto posto, Brooks fala sobre a arquitetura técnica, apresentando duas formas distintas (*engineering approaches*) de explicar o incremento da inteligência¹⁰²: a

¹⁰⁰ Brooks, “Intelligence Without Representation”, 140.

¹⁰¹ « I wish to build completely autonomous mobile agents that coexist in the world with humans, and are seen by those humans as intelligent beings in their own right. I will call such agents Creatures. » Rodney A. Brooks, “Intelligence without Representation”, *Artificial Intelligence* 47, n° 1–3 (janeiro de 1991): 142, [https://doi.org/10.1016/0004-3702\(91\)90053-M](https://doi.org/10.1016/0004-3702(91)90053-M).

¹⁰² De facto, o interesse de Brooks no desenvolvimento dos robôs é técnico e não exclusivamente filosófico – daí a abordagem técnica no modelo incremental da mente em seu texto. Brooks, 142.

decomposição por funcionalidade e a decomposição por atividade¹⁰³. A decomposição por funcionalidade¹⁰⁴ implica a separação das partes constitutivas do robô por módulos funcionais. Estes módulos são coordenados por um sistema central, que serve de interface às distintas funcionalidades: representação do conhecimento, aprendizado, planejamento, etc. O problema desta abordagem é a longa cadeia de módulos necessárias para que uma determinada ação possa ocorrer. Caso aconteça alguma falha técnica – também conhecida como *bug* – toda a cadeia responsável pela ação deixa de ser bem-sucedida e se torna ineficaz. Outro ponto crítico desta abordagem é a necessidade de prever todas as condicionantes necessárias para que tais módulos possam operar de forma eficaz, o que corrobora às objeções apontadas por Brooks ao uso das abstrações em robôs.¹⁰⁵

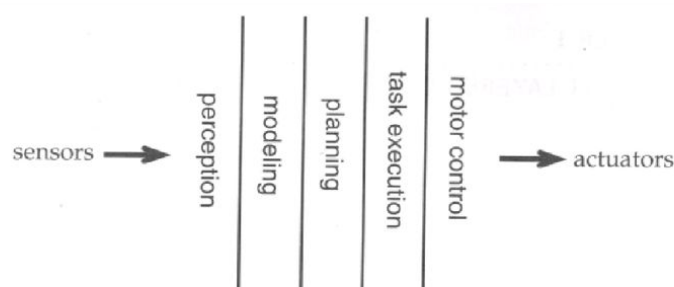


Figura 5 – Decomposição funcional ¹⁰⁶

Para elucidar mais detalhadamente o problema referente a abstração nos robôs, pode-se pensar na questão relacionada a percepção das cores. Entre os cientistas há divergências sobre se as cores realmente existem como uma propriedade do mundo ou se são meramente resultado dos sistemas visuais dos agentes. As abelhas, por exemplo, não percebem a cor vermelha da mesma forma que os seres humanos, apesar de possuírem sistema visual tricromático. Desta forma, elas percebem a radiação ultravioleta ao invés da radiação vermelha, uma característica que as auxilia na

¹⁰³ Brooks, 143.

¹⁰⁴ Ver Figura 5 – Decomposição funcional

¹⁰⁵ « When researchers working on a particular module get to choose both the inputs and the outputs that specify the module requirements, I believe there is little chance the work they do will fit into a complete intelligent system. » Rodney A. Brooks, “Intelligence without Representation”, *Artificial Intelligence* 47, nº 1–3 (janeiro de 1991): 143, [https://doi.org/10.1016/0004-3702\(91\)90053-M](https://doi.org/10.1016/0004-3702(91)90053-M).

¹⁰⁶ Rodney A. Brooks, “Achieving Artificial Intelligence through Building Robots”, AIM-899, 1º de maio de 1986, 7, <https://dspace.mit.edu/handle/1721.1/6451>.

navegação no mundo e na coleta de pólen.¹⁰⁷ Contudo, se fosse implementado um sistema visual em um robô abelha, qual seria a representação, ou abstração, que este agente deveria ter? Aquela percebida pela abelha ou pelo homem? A forma ideal de inteligência deveria simplesmente fornecer ao agente competências perceptivas/motoras que atuem de forma harmônica e autônoma – sem abstração e centralização de dados¹⁰⁸ – e deixar a cargo da agência a percepção particular do ambiente. Assim, ao invés de imputar representações do mundo aos robôs, deve-se permitir a eles perceberem o ambiente por si, tal como uma abelha, um tubarão e os demais agentes naturais que interagem com o mundo de maneiras distintas. Este é o padrão de inteligência que deve ser implementado nos robôs: sem o uso de RM, mas com autonomia perceptual/motora.

A segunda forma de incremento intelectual na abordagem técnica é a decomposição por atividades – ou habilidades. Neste modelo, as capacidades não são concebidas por módulos funcionais, mas sim por camadas (*layers*) independentes, cada qual responsável por uma determinada habilidade. Por exemplo: ao invés de implementar no robô um módulo funcional para responder a demandas sobre eventos físicos e espaciais em geral – o que exigiria do agente ter um grande volume de abstrações, regras e condicionais –, opta-se por adicionar uma camada na arquitetura do robô, a qual é responsável exclusivamente pela atividade de desvio de objetos.

De facto, a arquitetura de subsunção¹⁰⁹ permite a decomposição ortogonal e sofisticada das atividades. Assim, as habilidades recém implantadas são sempre adicionadas no topo da pilha de atividades já existentes na agência. Desse modo, cada camada tem um propósito específico, como por exemplo evitar que o robô esbarre em objetos. As camadas novas sempre conseguem se comunicar com as anteriores, e estas por sua vez permanecem sempre independentes das novas. Esta autossuficiência técnica é devida ao design elegante apresentado por Brooks, permitindo camadas

¹⁰⁷ Varela, Thompson, e Rosch, “The embodied mind”, 2016, 201–2.

¹⁰⁸ « There is no central locus of control. » Rodney A. Brooks, “Intelligence without Representation”, *Artificial Intelligence* 47, nº 1–3 (janeiro de 1991): 146, [https://doi.org/10.1016/0004-3702\(91\)90053-M](https://doi.org/10.1016/0004-3702(91)90053-M).

¹⁰⁹ Traduzido do termo inglês *subsumption architecture*. Este modelo de arquitetura permite a inclusão de novas camadas de atividades num contexto já implantado e complexo. Contudo, as novas camadas não interferem no funcionamento das anteriores, apesar de exercerem domínios sobre elas, garantindo a autonomia das atividades e evitando gargalos, como aqueles apontados na arquitetura de Von Newman, no programa do cognitivismo. Mais detalhes técnicos sobre a arquitetura: Brooks, 146.

independentes, cada qual constituída de sensores e ativadores, percebendo e respondendo aos estímulos do ambiente. Destarte, os robôs gozam de autonomia para interagir com o mundo, conforme as suas características percecionais e motoras.¹¹⁰

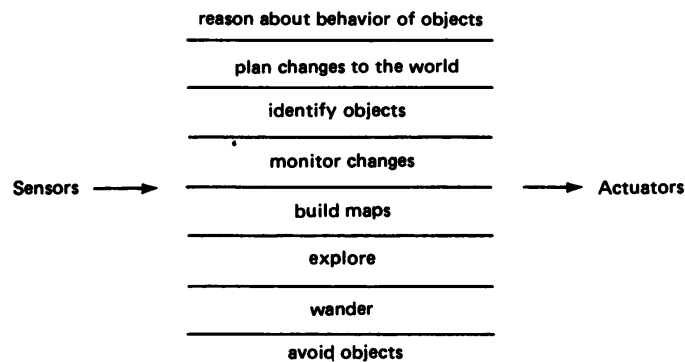


Figura 6 – Behavior-based decomposition.¹¹¹

Cada camada desta arquitetura é efetivamente composta por uma topologia fixa de máquinas de estados finitas (MEF), cada qual com os seus respectivos estados, um ou dois registros, um ou dois temporizadores e acesso a uma máquina computacional simples.¹¹² As máquinas de estados atuam de forma assíncrona, enviando e recebendo mensagens através de uma rede «data-driven»¹¹³ – sem controle central das mensagens, nem consequentemente uma central de RM.

¹¹⁰ « Each layer of control can be thought of as having its own implicit purpose (or goal if you insist). Since they are active layers, running in parallel and with access to sensors, they can monitor the environment and decide on the appropriateness of their goals. Sometimes goals can be abandoned when circumstances seem unpromising, and other times fortuitous circumstances can be taken advantage of. The key idea here is to be using the world as its own model and to continuously match the preconditions of each goal against the real world. » Rodney A. Brooks, “Intelligence without Representation”, *Artificial Intelligence* 47, nº 1–3 (janeiro de 1991): 144, [https://doi.org/10.1016/0004-3702\(91\)90053-M](https://doi.org/10.1016/0004-3702(91)90053-M).

¹¹¹ Brooks, “Achieving Artificial Intelligence through Building Robots”, 7.

¹¹² « The arrival of messages or the expiration of designated time periods cause the finite state machines to change state. The finite state machines have access to the contents of the messages and might output them, test them with a predicate and conditionally branch to a different state, or pass them to simple computation elements. There is no possibility of access to global data, nor of dynamically established communications links. There is thus no possibility of global control. All finite state machines are equal, yet at the same time they are prisoners of their fixed topology connections. » Brooks, 146.

¹¹³ Brooks, “Intelligence Without Representation”, 146.

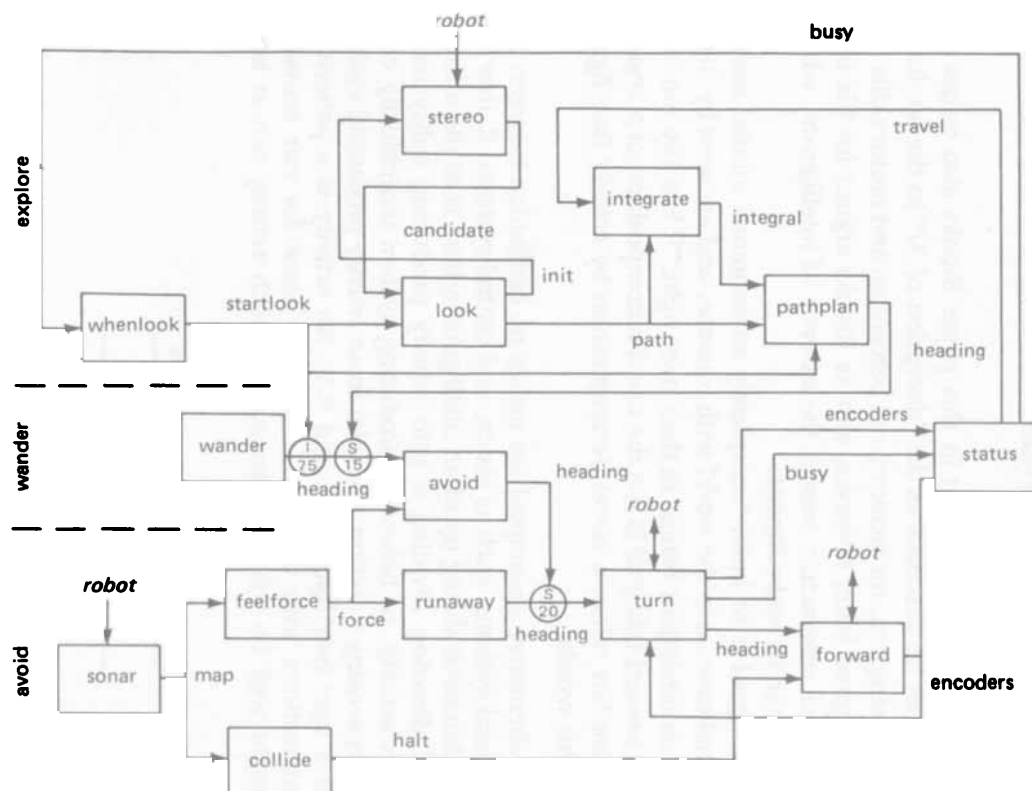


Figura 7 – Topologia das máquinas de estados finitas (MEF) dos robôs de Brooks¹¹⁴

3.1.1 A RM do tipo[Y] das criaturas de Brooks

Nesta seção será analisada o tipo de RM [Y] existente nas criaturas de Brooks, o que a princípio parece ser algo paradoxal, visto que a arquitetura de subsunção parece impedir o uso de representações informacionais. Contudo, averiguando minuciosamente o modelo arquitetural, nota-se que esta teoria foi delineada com uma topologia fixa de

¹¹⁴ « Finite state machines are wired together into layers of control. Each layer is built on top of existing layers. Lower levels never rely on the existence of higher-level layers. » Rodney A. Brooks, “Achieving Artificial Intelligence through Building Robots”, AIM-899, 1º de maio de 1986, 8, <https://dspace.mit.edu/handle/1721.1/6451>.

máquinas de estados finitas (MEF). Cada MEF é realmente um modelo computacional que simula sequências lógicas e alguns programas de computador, podendo ser implementada em hardware – num torniquete, por exemplo – ou em software – num sistema de informação. As MEF geralmente são usadas para solucionar problemas de matemática, inteligência artificial, jogos e linguística¹¹⁵.

No caso do torniquete, que é um exemplo de MEF implementada no hardware¹¹⁶, pode-se descrever a dinâmica da máquina da seguinte forma: o estado inicial do torniquete é *Locked*; caso o agente insira uma moeda o estado da máquina é atualizado para *Unlocked*. Ao empurrar o torniquete o estado novamente é atualizado para *Locked*. A dinâmica será sempre esta, não sendo possível ao agente adicionar uma moeda no estado *Locked*, empurrar o torniquete e não ocorrer a transição para o estado *Unlocked*. As exceções seriam a ocorrência de alguma falha mecânica ou o repositório de moedas estar completo. A MEF do torniquete tem este funcionamento porque o engenheiro responsável pela estrutura deste autômato criou-o com esta dinâmica, simples assim. Portanto, o modelo a ser seguido é aquele que está constituído na estrutura do torniquete.¹¹⁷

Voltando à arquitetura de subsunção de Brooks, constata-se que as suas criaturas estão determinadas estruturalmente do mesmo modo que o torniquete supracitado.

All finite state machines are equal, yet at the same time they are prisoners of their fixed topology connections.¹¹⁸

¹¹⁵ “Finite State Machines | Brilliant Math & Science Wiki”, acessado 2 de maio de 2019, <https://brilliant.org/wiki/finite-state-machines/>.

¹¹⁶ Cfr. Figura 8 – MEF de um torniquete, pág. 66

¹¹⁷ “Finite State Machines | Brilliant Math & Science Wiki”.

¹¹⁸ Brooks, “Intelligence Without Representation”, 146.

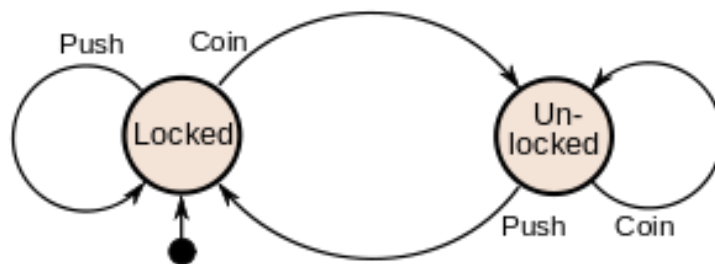


Figura 8 – MEF de um torniquete ¹¹⁹

Isto pode ser comprovado analisando a arquitetura de subsunção constitutiva das três camadas de atividades apresentadas acima: *Avoid*, *Wander* e *Explore*.¹²⁰ A camada *Avoid* é composta de diversas MEF, dentre as quais duas *estão* conectadas diretamente ao *Sonar*: *Feel Force* e *Colide*. O *Sonar* é um anel de doze sonares ultrassônicos que mapeia e envia coordenadas simultaneamente para as MEF conectadas. A MEF *Feel Force* usará estas coordenadas para calcular a força repulsiva que será usada na MEF subsequente: *Runaway*. Já a MEF *Colide* verifica se há algum objeto parado a frente para que não ocorra nenhuma colisão.¹²¹ Dessa forma, as MEF atuaram sempre com a mesma dinâmica, ou em outros termos, os robôs terão sempre o mesmo desempenho nesta camada, não havendo possibilidade de, por exemplo, a estrutura operar primeiro a MEF *Colide*, para posteriormente acessar a máquina *Feel Force*.

Dessa forma, uma MEF não deixa de ser um modelo representativo, um mapa conceitual de um universo de estados determinados, e conseqüentemente, também uma forma de abstração. A própria topologia das camadas é um mapa, uma abstração posicional e distribuída de elementos físicos, idealizada e montada por engenheiros, afim de que o robô obtenha a forma mais eficiente e eficaz de atuação no ambiente. Portanto, as RM das criaturas de Brooks não estão em programas, banco de dados, condicionas e algoritmos, estão em suas estruturas físicas – ou, conforme também se pode dizer, as criaturas de Brooks possuem RM do tipo *embodied structures* ou RM estruturais.

¹¹⁹ “Finite State Machines | Brilliant Math & Science Wiki”.

¹²⁰ Cf. Figura 7 – Topologia das máquinas de estados finitas (MEF) dos robôs de Brooks 64.

¹²¹ Brooks, “Intelligence Without Representation”, 146.

Atendendo à arquitetura de subsunção das criaturas de Brooks, consideramos como estrutura para a RM do tipo [4], ou enativismo implementacional, as seguintes condições:

- I. A estrutura das habilidades dos agentes é determinada pela arquitetura de subsunção;
- II. A topologia estrutural dos agentes não apresenta plasticidade, é fixa;
- III. A atuação é resultado das interações das camadas de atividades definidas na topologia estrutural;
- IV. Não há posse de RM descritivas, regras ou condicionais;
- V. Como a estrutura física dos agentes é a implementação de abstrações conceituais e modelos de atuação, o enativismo implementacional também se caracteriza como uma forma de RM.

Por fim, definida as condições da RM do tipo enativismo implementacional¹²² no qual estão enquadrados os robôs de Brooks, resta considerá-las no âmbito do experimento do bastão.

3.1.2 As criaturas de Brooks e o experimento do bastão

As criaturas de Brooks mostram que é possível ao robô apresentar inteligência. Assim, este tipo de robô é capaz de interagir de forma inteligente em um ambiente dinâmico, e aqui entenda-se inteligência como capacidade de interpretar o ambiente e transpor as adversidades apresentadas por ele. Empecilhos, como o desvio de uma cadeira, são possíveis devido ao modelo de arquitetura de subsunção¹²³ que permite a inserção de habilidades de forma incremental em estruturas complexas existentes. Contudo, isto atende às demandas cognitivas do experimento do bastão?

De acordo com as condicionantes descritas para as RM do tipo [4], ou enativismo implementacional, é possível afirmar que o agente portador deste tipo de RM não estaria

¹²² Esta tese sempre usará o termo «implementacional» para designar as implementações particulares das RM do tipo [Y] propostas pelos autores antirrepresentacionistas. Logo, um RM do tipo [Y] implementacional é uma variante particular dos principais programas cognitivistas, mas constituídas de alguma forma em sua estrutura de RM. Neste caso específico, a estrutura dos robôs de Brooks usam muitos dos princípios enativistas, portanto, este tipo de RM será designado de RM enativista implementacional.

¹²³ Brooks, “Intelligence without Representation”, 146.

apto ao cumprimento da condicionante ET4, – Incremento cognitivo, definido no experimento do bastão¹²⁴. Supondo um cenário no qual o robô apresente três camadas de habilidades – locomoção, procurar frutas e geolocalização –, como seria possível a criatura brookiana aprender ou acrescentar uma nova camada para a habilidade «construir bastão»? A única forma de contornar esta obstáculo seria permitir ao agente atualizar a arquitetura de subsunção (I), adicionado uma nova camada na estrutura existente e permitindo ao robô satisfazer a demanda explicitada no experimento.

Todavia esta atualização violaria as restrições impostas pelo experimento, nomeadamente a de não ocorrer nenhuma forma de atualização e interferência técnica/estrutural nos agentes. As topologias fixas dos robôs (II) não lhes permitem aprender novas habilidades, e assim como a MEF do torniquete, os robôs de Brooks estariam condenados a dinâmica representacional incrustada em suas estruturas, até o momento que algum agente externo resolvesse fazer alguma intervenção/atualização. Alguém pode apontar, contudo, para o fato do texto de Brooks ter sido escrito há 32 anos e alguma modificação poder ter ocorrido neste período.

Uma forma de averiguar esta questão é avaliar o histórico dos projetos nos quais o autor esteve envolvido. Começando pela empresa *iRobot*: a empresa continua ativa no mercado e os robôs residenciais são amplamente populares. O modelo conhecido como *Roomba 960*¹²⁵ é capaz de limpar tapetes e superfícies, inclusive em lugares de difícil acesso, como por exemplo, debaixo de uma cama. Também é possível programar os horários de limpeza do robô via aplicação celular. Assim, o robô tem os atributos cognitivos de III – ter experiência de; IV – conhecer pela experiência; V – tornar-se ciente e VI – discernir¹²⁶, dentro das especificações que lhe foram atribuídas.

Porém, o *Roomba 960* não possui os atributos que possam transformar esta atuação em cognição. O modelo *Roomba 960* não consegue, por exemplo, «I – aprender» que o melhor horário de limpar a casa é durante as manhãs, quando todos os moradores estão ausentes, ou «II – descobrir» que a sala é o cômodo mais desfrutado

¹²⁴ Cfr. seção 2.2 O experimento do bastão, 46.

¹²⁵ iRobot, “Robô Aspirador de Pó Inteligente Roomba 960 iRobot”, acessado 1º de maio de 2019, <https://www.irobotloja.com.br/roomba-960-robo-aspirador-de-po-inteligente-bivolt-irobot-r960400/p>.

¹²⁶ Cfr. seção Observações sobre palavras-chave, 16.

pelos moradores, e, portanto, merece mais atenção no processo de limpeza; ou «VII – reconhecer» os moradores da casa desejando-lhes bom dia ou chamando-os pelos nomes.¹²⁷

A outra empresa em que Brooks esteve envolvido foi a *Rethink Robotics*, que recentemente foi adquirida pelo grupo alemão HAHN. O objetivo de Brooks foi criar robôs que pudessem ser usados em linhas de produção de forma descomplicada, na execução de tarefas rotineiras: pegar objetos, empilhá-los, colocá-los em esteiras, e demais tarefas de linhas de produção. Os robôs seguiam as diretrizes do que foi apresentado até o momento, ou seja, não há programas centrais, mas são já capazes de desenvolver alguma forma de cognição, sendo assim modelos mais sofisticados do que o *Roomba 960*. Contudo, a evolução cognitiva dos robôs *Baxter*¹²⁸ – o modelo da *Rethink* para linhas de produção – mostrou-se muito mais difícil do que se esperava, resultando em baixos números de vendas e desempenho. Pelo que, considerando o histórico dos dois empreendimentos e o texto de Brooks, pode-se constatar que estruturas autônomas não são capazes, só por isso, de atender de forma satisfatória a autonomia cognitiva da agência.

Worldwide, manufacturers are trying to automate ever more tasks to make up for shortages of human labor. E-commerce giants, meanwhile, are exploring new approaches to automating the picking, packing, and processing of goods. There's a catch, though. If you visit a company using robots, you'll find that the vast majority aren't very smart or adaptive. They can do things precisely and tirelessly, but they are easily befuddled by real-world complexity and are mostly painful to program.¹²⁹

Estes fatos podem ajudar a explicar o crescente interesse em pesquisas de *deep learning*, o que retorna a discussão para a necessidade de algum tipo de RM para a autonomia cognitiva em sistemas de IA, contrariamente ao que é o propósito de Brooks. Pode-se, pois, afirmar que as condicionantes impostas a cognição nesta tese, I –

¹²⁷ Cfr. seção Observações sobre palavras-chave, 16.

¹²⁸ “Sawyer Collaborative Robots for Industrial Automation”, *Rethink Robotics*, acessado 3 de maio de 2019, <https://www.rethinkrobotics.com/sawyer/>.

¹²⁹ Will Knight, “The Demise of Rethink Robotics Shows How Hard It Is to Make Machines Truly Smart”, *MIT Technology Review*, 4 de outubro de 2018, 16.

aprender, II – descobrir e VII – reconhecer algo ou alguém já conhecido¹³⁰, ainda não foram solucionadas pelos empreendimentos herdeiros das criaturas de Brooks. Num cenário mais otimista, os robôs continuam com a capacidade cognitiva equivalente a um inseto.¹³¹

Assim, a cognição continua a ser um problema de árdua e complexa resolução, conforme apontou a cientista Manuela Veloso¹³² na conferência “Fatores para o sucesso da interação com a Inteligência Artificial.”¹³³ Essa pesquisadora salienta que devido as limitações cognitivas em sistemas de IA, a próxima fase desta ciência deverá ser a de interação e simbiose com os humanos. Deste modo, por exemplo, se um robô não consegue apertar o botão de um elevador, ele deverá ser capaz de pedir a ajuda a algum humano para fazê-lo. Ainda assim, Veloso, tal como Brooks, salientam a importância da autonomia dos robôs, com características motoras e perceptuais que lhes permitam interagir no mundo. Ambos os autores acreditam que este é o caminho para fazer evoluir a IA.

Mas percebe-se também uma tendência entre os roboticistas para admitir os limites dos sistemas de IA, e daí que termos como «simbiose» e «colaboração» estejam cada vez mais presentes na área da IA. Mesmo em projetos de robótica altamente sofisticados, como o robô *Cheetah III*, um robô de quatro pernas que executa os mesmos movimentos do felino homônimo, as diretrizes de construção pensam na cooperação com os humanos. Este é o posicionamento do professor de engenharia

¹³⁰ Cfr. seção Observações sobre palavras-chave, 16.

¹³¹ « *We believe they operate at a level closer to simple insect level intelligence than to bacteria level intelligence.* » Rodney A. Brooks, “Intelligence without Representation”, *Artificial Intelligence* 47, nº 1–3 (janeiro de 1991): 148, [https://doi.org/10.1016/0004-3702\(91\)90053-M](https://doi.org/10.1016/0004-3702(91)90053-M).

¹³² Manuela M. Veloso integrou recentemente a J.P.Morgan Chase para criar e liderar um Centro de Investigação em Inteligência Artificial (IA). Veloso está de licença da Universidade Carnegie Mellon (CMU), onde é Professora da Cátedra Herbert A. Simon da Escola de Ciência da Computação, e onde foi chefe do departamento de Machine-Learning até junho de 2018. A sua investigação incide nas áreas de IA, Robótica e Machine-Learning. Na CMU, fundou e dirige o laboratório de investigação CORAL, para o estudo de agentes autônomos que colaboram, observam, raciocinam, agem e aprendem. Veloso e os seus estudantes investigam uma variedade de robôs autônomos, incluindo robôs de serviço móvel e robôs de futebol. Veloso é “Fellow” das associações profissionais e académicas da sua área de investigação, nomeadamente AAAI, ACM, AAAS e IEEE. É co-fundadora e ex-presidente do RoboCup e ex-presidente do AAAI.

¹³³ Manuela Veloso, “Fatores para o Sucesso da Interação com a Inteligência Artificial” (Premium Conference, Universidade Católica Portuguesa, 2018).

mecânica do MIT Sangbae Kim¹³⁴: « *This is my big vision: human-level mobility, mostly autonomous, with the manipulation mostly done by humans.* »¹³⁵

Dito isto, a que conclusões se pode chegar? Estes fatos corroboram a tese de que a posse de RM do tipo presente no enativismo implementacional não basta para a autonomia cognitiva da agência. Assim, a cognição não seria possível sem algum tipo de RM. Contudo, Brooks foi sagaz ao apontar que atualmente os sistemas de IA não têm passado de modelos reduzidos das abstrações feitas pela mente humana. Seria preciso encontrar antes de tudo uma forma dos robôs experimentarem o mundo com autonomia, mas sem a imposição de estruturas fixas que lhes restrinjam o desempenho cognitivo. Isto reforça a necessidade de investigar a natureza da flexibilidade e adaptabilidade do processamento das RM na agência humana para que, quiçá, elas possam ser transpostas para sistemas de IA. Pode-se dizer ainda que Brooks estava equivocado em sua interpretação sobre a evolução biológica da inteligência, pois, ao que parece, a relação entre características físicas – percecionais/motoras – e inteligência não são afinal tão óbvias quanto ele acreditava. Isto explicaria a discrepância entre projetos robóticos como *Cheetah III* e *Baxter*¹³⁶ e os ambicionados progressos relacionados a cognição em sistemas de IA.

3.2 Pengi: uma implementação da teoria da atividade

Até esta altura a posse e o processamento das RM parecem ser condições necessárias para a cognição em sistemas de IA. Todavia, no texto *Pengi: an implementation of a theory of activity*¹³⁷, os filósofos Philip P. Agre e David Chapman contestam a necessidade das RM em tarefas rotineiras, visto que a interação entre os agentes e o ambiente é tão dinâmica e complexa, que se torna impossível de ser

¹³⁴ “MECHE PEOPLE: Sangbae Kim | MIT Department of Mechanical Engineering”, acessado 3 de maio de 2019, <http://meche.mit.edu/people/faculty/SANGBAE@MIT.EDU>.

¹³⁵ Cara Giaimo, “Sangbae Kim Ditched the Rules of Robotics and Built Tomorrow’s first Responder.”, [s.d.], 18.

¹³⁶ « *I believe that mobility, acute vision and the ability to carry out survival related tasks in a dynamic environment provide a necessary basis for the development of true intelligence.* » Rodney A. Brooks, “Intelligence without Representation”, *Artificial Intelligence* 47, nº 1-3 (janeiro de 1991): 140, [https://doi.org/10.1016/0004-3702\(91\)90053-M](https://doi.org/10.1016/0004-3702(91)90053-M).

¹³⁷ Agre e Chapman, “Pengi”, 1987.

planejada a todos os instantes. Assim, a maior parte das respostas aos desafios da agência no mundo ocorreriam de forma improvisada, sem RM associadas a planos preconcebidos.¹³⁸

Os autores começam por considerar que as ações dos agentes podem ser divididas em dois grupos de atividades: planejamento com «P» e planejamento com «p»¹³⁹. O planejamento com «P» é a aquele que necessita de um tempo para ser executado, como por exemplo, a construção do bastão por Hayy.¹⁴⁰ Contudo, na maior parte das vezes, Hayy precisou lidar com decisões imediatas, como subir e descer de árvores, procurar comida, beber água, abrigar-se do frio e outras tarefas que demandaram ações rápidas, nas quais não há tempo para o planejamento «P». Para estas tarefas rotineiras, Chapman e Agre definem o planejamento «p».¹⁴¹

O planejamento «P» é operacionalmente inviável na maioria das necessidades da agência, pois não há tempo, nem recursos e informações suficientes para lidar exaustivamente com as situações reais da vida que emergem a todos os instantes. Afinal, seria impraticável planejar com «P» a cada momento as tarefas do cotidiano, como ir ao trabalho, descascar batatas ou desviar de um buraco do chão. Assim, uma boa parte das ações dos agentes são decorrentes de mecanismos imediatos.¹⁴²

Para comprovar esta teoria da atividade, os autores reimplementaram o jogo Pengo – um famoso jogo Arcade dos anos oitenta¹⁴³ – renomeando-o para Pengi. O objetivo foi simular a complexidade do mundo através da dinâmica contextual do próprio jogo, o qual é constituído de um pinguim, um labirinto de blocos de gelo e algumas abelhas caçadoras. A dinâmica consiste de as abelhas caçarem o pinguim antes

¹³⁸ « *Rather than relying on reasoning to intervene between perception and action, we believe activity mostly derives from very simple sorts of machinery interacting with the immediate situation.* » Philip E. Agre e David Chapman, “Pengi: An Implementation of a Theory of Activity”, 1987, 268.

¹³⁹ Traduzido dos termos em inglês: *capital-P planning* e *lower-case-p planning*. Agre e Chapman, 268.

¹⁴⁰ Cfr. seção 2.1 A primeira fase cognitiva de Hayy, 44.

¹⁴¹ Agre e Chapman, “Pengi”, 1987, 1.

¹⁴² « *Planning is inherently combinatorially explosive, and so is unlikely to scale up to realistic situations which take thousands of propositions to represent. Most real situations cannot be completely represented; there isn't time to collect all the necessary information. Real situations are rife with uncertainty; the actions of other agents and processes cannot be predicted.* » Agre e Chapman, 268.

¹⁴³ “Pengo (Video Game)”, in *Wikipedia*, 14 de março de 2019, [https://en.wikipedia.org/w/index.php?title=Pengo_\(video_game\)&oldid=887735053](https://en.wikipedia.org/w/index.php?title=Pengo_(video_game)&oldid=887735053).

que ele possa acionar os blocos mágicos contidos no labirinto. Além disto, tanto as abelhas quanto o pinguim podem chutar blocos de gelos uns contra os outros e eliminarem-se do jogo. Desta forma, o ambiente do jogo permite a profusão de possibilidades que são muito difíceis de prever, tornando possível simular – dadas as devidas proporções – o mundo real. A reimplementação de Pengi seria assim um valioso laboratório para investigar a interação e a cognição da agência via rotinas.¹⁴⁴

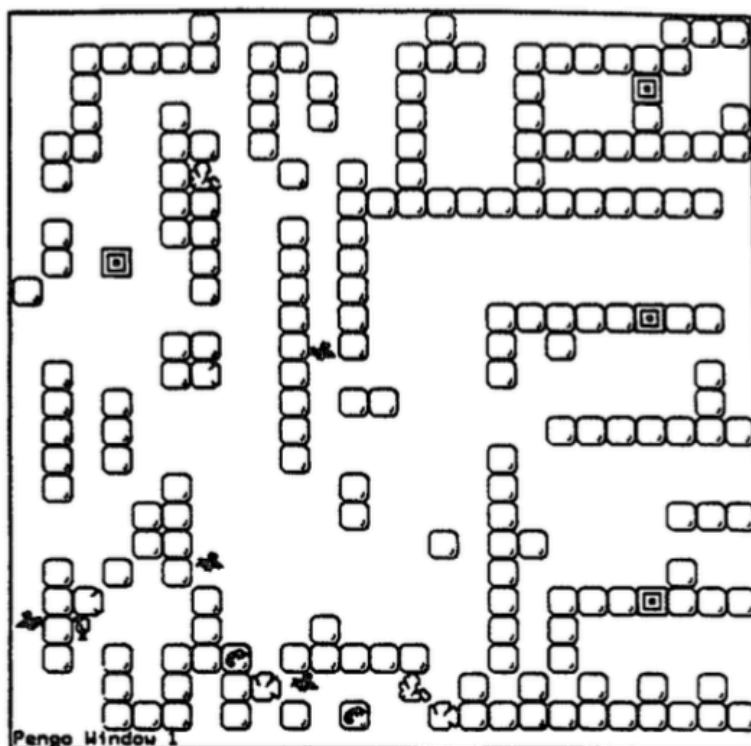


Figura 9 – Tela do jogo Pengo.¹⁴⁵

Com relação as rotinas, estas são definidas pelos autores como padrões de interação entre o agente e o seu mundo¹⁴⁶, permitindo-lhes tratar dos imprevistos e das oportunidades emergentes no ambiente, sem a necessidade de roteiros preconcebidos ou

¹⁴⁴ Além do que foi exposto sobre o jogo, acrescenta-se que os autores imputaram objetivos e habilidades nos personagens para que o jogo Pengo pudesse ser jogado pelo sistema de IA.

¹⁴⁵ Agre e Chapman, “Pengi: An Implementation of a Theory of Activity”, 269.

¹⁴⁶ Agre e Chapman, 269.

planejamento «P». De maneira oposta, os roteiros preconcebidos e procedimentos¹⁴⁷ são funções constituídas de instruções planejadas para solucionar apenas problemas específicos. Assim, os procedimentos exigem planejamento prévio do roteiro que será executado pela agência na resolução dos problemas. Neste modelo de atividade, a atuação dos agentes é baseada na iteração integral, sem a possibilidade de flexibilizar a execução das instruções contidas no procedimento. Por fim, pode-se dizer ainda que as rotinas se caracterizam como padrões de interação entre o agente e o seu ambiente e de maneira oposta, os procedimentos por regras preconcebidas de atuação¹⁴⁸, que implicam a necessidade de manter planos e RM para a atuação.

Como se configuraria a ação em Pengi, a partir de uma rotina e de um procedimento, dado o evento «pinguim foge das abelhas»?

Com relação ao procedimento, este seria composto por instruções que deveriam ser executadas obrigatoriamente de forma sequencial:

- i. correr até encontrar um bloco de gelo;
- ii. chutar o bloco de gelo contra a abelha;
- iii. repetir procedimento, caso o evento «pinguim foge das abelhas» continue.

Note-se que os procedimentos são condicionados à iteração (iii), ou em outras palavras, a recursividade integral do plano. Assim, sempre que o evento «pinguim foge das abelhas» for evocado, o mesmo roteiro de instruções será executado. Além disto, há a questão relacionada com a inflexibilidade de ativação das instruções de forma atômica, como por exemplo, executar apenas a instrução (ii).

Por sua vez, as rotinas apresentam características mais flexíveis, pois uma rotina «pinguim foge das abelhas» seria delineada pelo padrão da interação entre o agente e o seu mundo e não por um plano preconcebido, como no caso dos procedimentos. Além disto, é permitido ao agente flexibilizar a execução das instruções, podendo atuá-las em grupo ou de forma atômica, de acordo com a situação na qual se encontre. Portanto, as

¹⁴⁷ Tradução do termo inglês: *Procedures*. *Procedures*, é um conjunto de instruções usadas como rotinas e funções em banco de dados. Esses conjuntos de instruções são usados em tarefas repetitivas de dados – *INSERT*, *UPDATE*, *DELETE* – e são armazenadas em servidores, encapsulando muitas vezes as regras de negócio de uma empresa.

¹⁴⁸ Agre e Chapman, “Pengi: An Implementation of a Theory of Activity”, 268.

rotinas são oportunistas, ativando-se de acordo com a particularidade da situação, e são robustas quanto à incerteza dos momentos, liberando a agência da imposição de possuir roteiros e RM. Poder-se-ia pôr como exemplo de rotina o seguinte:

- i. correr até encontrar um bloco de gelo;
- ii. chutar o bloco de gelo contra a abelha.

A penguin controlled by an iterative procedure would then either continue running needlessly or have to notice that it had gone wrong and switch to executing a different procedure. An agent engaging in a routine is not driven by a preconceived notion of what will happen.¹⁴⁹

Desse modo, seria possível ativar as instruções de forma atômica – (i) ou (ii) – ou ativá-las de forma sequencial – (i) e (ii). Tendo isto em vista é possível afirmar que as rotinas atendem a autonomia cognitiva em sistemas de IA? Para responder esta questão é necessário imergir na arquitetura de Pengi.

Para averiguar a dinâmica da teoria da atividade de Chapman e Agre, considere-se como exemplo a situação mais básica que pode haver no jogo: uma abelha qualquer perseguindo o pinguim. Deste ponto de vista, as abelhas, o pinguim e os blocos de gelo são sistemas periféricos do jogo, desprovidos de procedimentos relativos a como jogá-lo¹⁵⁰. A estrutura destes sistemas é constituída pelo sistema perceptual, responsável pela visão, e pelo sistema motor, responsável por atuar no ambiente. Todavia, o gerenciamento da ação não é conduzido pelos sistemas periféricos, mas por rotinas

¹⁴⁹ Agre e Chapman, 269.

¹⁵⁰ Não há uma indicação explícita no texto de que os pinguins, as abelhas e os blocos sejam sistemas periféricos. Esta ideia é o posicionamento da tese a partir de referências ao longo do texto: «sistemas periféricos» ao invés de «sistema periférico»; «sistemas perceptuais» ao invés de «sistema perceptual» e «sistemas de controle de motores» ao invés de «sistema de controle de motor». Agre e Chapman, seq. IV. Simple Machinery, 270.

visuais operantes entre o sistema central¹⁵¹ e o componente denominado de processador visual de rotinas (PVR).

A dinâmica entre os sistemas perceptuais e o PVR orchestra-se da seguinte forma: processos contínuos de rotinas visuais copiam os esboços 2D produzidos pelos sistemas perceptuais (*early vision*¹⁵²) para o PVR. Esses esboços são imagens das situações particulares vivenciadas pelos agentes, como no exemplo citado, a imagem representante da situação do pinguim fugindo da abelha nos corredores do labirinto de gelo. Esta cópia permitirá ao componente PVR operar diversas funções sobre a imagem, tais como colorir regiões, rastrear curvas, manter rastreamento de localizações usando marcadores nas imagens, indexar características interessantes e detetar/rastrear objetos em movimento.¹⁵³

Além de operar sobre as imagens, o PVR também é responsável por procurar as entidades e registrar os aspectos das situações enviadas para ele. Mas o que são entidades e aspectos?

As entidades ou entidades funcionais indexicais¹⁵⁴ são elementos intermediários entre os indivíduos lógicos e as categorias¹⁵⁵ do jogo. Com relação as propriedades das entidades, a indexicalidade refere-se a circunstância vivida na agência e a funcionalidade aos objetivos e motivos dela. Dito em outras palavras, as entidades são

¹⁵¹ Os autores não definem explicitamente o que é o sistema central. Primeiro há uma distinção entre sistema central e sistemas periféricos. « *This architecture is made up of a central system and peripheral systems.* » Agre e Chapman, 270. Depois, uma distinção entre sistema central e o processador visual de rotinas (PVR): « *visual routines which are patterns of interaction between the central system and a visual routines processor (VRP).* » Agre e Chapman, 270. E por fim há uma consideração de que as redes se combinam, adequam e formam um apropriado sistema central. « *We believe that combinational networks can form an adequate central system for most activity.* » Agre e Chapman, 270. Portanto, neste capítulo quando se usar o termo sistema central, estar-se-á referindo a rede combinacional constituída pelos sistemas periféricos. Contudo, o sistema central pode conter outras funcionalidades, distintas do PVR e dos sistemas periféricos. Ver Anexo 6 – Diagrama genérico de um circuito combinacional.

¹⁵² Agre e Chapman, “Pengi”, 1987, 268.

¹⁵³ Agre e Chapman, 271.

¹⁵⁴ O termo indexical, refere-se as expressões linguísticas que variam o sentido de acordo com o contexto, por exemplo a palavra «aqui». As entidades e os aspectos são indexicais porque precisam do seu referencial – situação particular – para fazer sentido.

¹⁵⁵ Chapman e Agre não citam no texto exemplos de indivíduos lógicos e categorias. O autor Stan Franklin, que analisa o respectivo texto, coloca como exemplos de tais elementos: « *Entities are intermediate between logical individuals and categories. BLOCK-213 is a logical individual. The set of blocks being pushed is a category. The-block-I'm-pushing is neither but something in between.* » Stan Franklin, *Artificial Minds*, A Bradford Book (Cambridge, Mass.: MIT Press, 2001), 372.

representações indexicais e funcionais das circunstâncias e dos propósitos dos agentes, respectivamente. São exemplos de entidades:

- *the-block-I'm-pushing*
- *the-corridor-I'm-running-along*
- *the-bee-on-the-other-side-of-this-block-next-to-me*¹⁵⁶

Por sua vez, os aspectos ou aspectos funcionais indexicais¹⁵⁷ são tipos particulares de entidades, nas quais estão impressas as propriedades da situação corrente em que o agente está envolvido. São exemplos de aspectos:

- *the-block-I'm-going-to-kick-at-the-bee-is-behind-me* (so I have to backtrack)
- *there-is-no-block-suited-for-kicking-at-the-bee* (so just hang out until the situation improves)
- *I've-run-into-the-edge-of-the-screen* (better turn and run along it)¹⁵⁸

Outro ponto que deve ser ressaltado é que os aspetos não estão vinculados a individualidades, ou seja, o aspecto *the-bee-that-is-chasing-me*, não se refere a uma instância específica de abelha, mas a qualquer abelha posta em situação idêntica. Mas por que esta característica é importante? Porque permite ao Pengi rodar o jogo Pengo sem o artifício de RM específicas para manter o estado das unidades lógicas, além de tecnicamente ser mais rápido do que o modelo de instâncias, importante quando se operam atividades em *real-time*¹⁵⁹. Assim, o carácter indexical dos aspectos permite que não seja necessário lidar com instâncias de abelhas, pinguins e blocos, e consequentemente evita o peso de também representar as propriedades destes objetos individuais. Deste modo desfaz-se boa parte da necessidade de representação do mundo nos sistemas de IA. Em outras palavras, manter estados de sistemas – com RM – é menos importante do que parece.

¹⁵⁶ Agre e Chapman, “Pengi”, 1987, 269.

¹⁵⁷ As definições das propriedades indexicais e funcionais nas entidades também se aplicam aos aspectos.

¹⁵⁸ Agre e Chapman, “Pengi”, 1987, 270.

¹⁵⁹ « *Avoiding the representation of individuals bypasses the overhead of instantiation: binding constants to variables.* » Agre e Chapman, 270.

We should comment, though, on Pengi's central system's lack of state. We do not believe that the human central system has no state. Our point is simply that state is less necessary and less important than is often assumed.¹⁶⁰

Ainda com relação aos aspectos, saliente-se que cada um é definido para uma resposta específica. Por isto, devem ser registrados vários aspectos *the-bee-on-the-other-side-of-this-block-next-to-me*, pois cada um pode se referir a uma situação de oportunismo para matar uma abelha ou de perigo exigindo a fuga imediata. Haverá assim dois aspectos *the-bee-on-the-other-side-of-this-block-next-to-me*, cada qual forjado a partir de uma circunstância do jogo.

Voltando ao componente PVR, agora a par dos conceitos de entidades e aspectos, torna-se possível continuar a esmiuçar o mecanismo arquitetural em Pengi. Assim, as rotinas visuais orquestram os processos de envio dos esboços 2D das circunstâncias vivenciadas pelos sistemas periféricos para o PVR. A partir daí, o PVR procura as entidades contidas nestes esboços, realiza operações sobre tais imagens – quando necessário – e registra os aspectos. Por fim, os aspectos são enviados para os sistemas motores dos sistemas periféricos, resultando na performance das ações.

Na Figura 10 é possível apreciar mais detalhadamente a dinâmica entre o sistema periférico e o PVR, a partir da situação na qual o pinguim está a espreitar a abelha que se encontra do outro lado do muro. Conforme se nota no exemplo, o PVR cadastrou o aspecto *the-block-to-kick-at-the-bee* através da situação enviada pelos sistemas perceptuais – esboço 2D. Na sequência, o PVR opera sobre a imagem e traça uma linha perpendicular entre o muro que está a frente do pinguim e a abelha que está do outro lado, encontrando através da intersecção, o ponto ou, melhor, o bloco que deverá ser chutado contra a abelha. A seguir será gerada uma saída, associada ao aspecto *the-block-to-kick-at-the-bee* e, por fim, enviado para uma rede, que se encarregará da atuação, devido as centenas de portas lógicas que compõem essa rede combinacional.¹⁶¹

¹⁶⁰ Agre e Chapman, 272.

¹⁶¹ Agre e Chapman, 272.

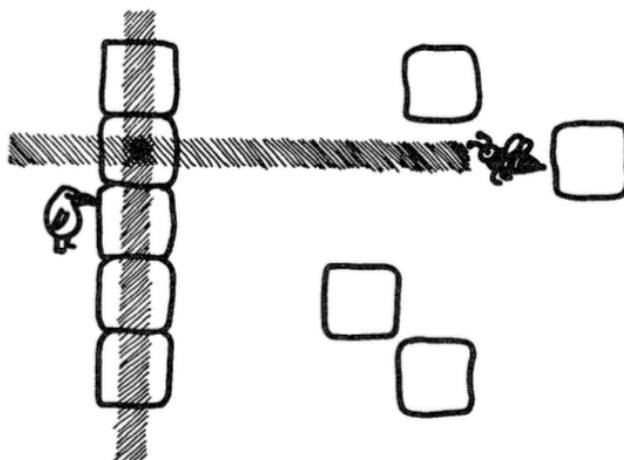


Figura 10 – Finding the-block-to-kick-at-the-bee when lurking behind a wall.¹⁶²

À vista disto, torna-se possível para o agente despojar-se do artifício das RM para a ação, e por sinal, ele não se importa e nem ao menos sabe disto. As ações como fugir ou chutar blocos de gelo serão executadas pelos sistemas motores como processos resultantes dos aspectos e isto é tudo o que basta para o desempenho dos agentes. A agência é liberada da posse e da armazenagem de procedimentos, variáveis, estados, bem como a necessidade de consulta ao bancos de dados.¹⁶³ Tudo o que Pengi precisa para operar o jogo é olhar para a tela – eis aqui a rotina –, e o resto do mecanismo é orquestrado pelo modelo arquitetural, delineado propositadamente para a interação das unidade lógicas com o ambiente ao invés da representação de modelos do jogo nelas.

Isto posto, surge a questão: quem gere afinal as atividades em Pengi? A gestão é compartilhada parte pelo sistema central e parte pelos sistemas periféricos. É importante

¹⁶² Agre e Chapman, 271.

¹⁶³ « Aspects, like routines, are not data structures. They do not involve variables bound to symbols that represent objects. Aspects are registered by routines in which the network interacts with the perceptual systems and with the world. The actions in these routines get the world and the peripheral systems. » Agre e Chapman, 270.

salientar que esses sistemas compõem uma rede combinacional¹⁶⁴, que se incumbe da maioria das atividades do jogo. Assim, o fluxo informacional da rede é retroalimentado a partir das entradas e saídas dos sistemas periféricos e dos respectivos registros no PVR. Dito em outras palavras, as iterações das interações circunstanciais – a recursividade das rotinas – entre os sistemas periféricos, o PVR é orquestrado por meio das rotinas visuais que forjam o ambiente do jogo e também são forjados por ele. Notam-se assim algumas semelhanças com as estruturas dos tipos de RM enativista¹⁶⁵ e o com o tipo de RM subsimbólico¹⁶⁶, pois a capacidade de interação dos agentes no ambiente do jogo depende das estruturas desses agentes e essa interação ocorre também de forma distribuída.

We believe that combinational networks can form an adequate central system for most activity. The inputs to the combinational network come from perceptual systems; the outputs go to motor control systems. The network decides on actions that are appropriate given the situation it is presented with. Many nodes of the network register particular aspects. As the world changes, the outputs of the perceptual system change; these changes are propagated through the network to result in different actions. Thus interaction can result without Pengi maintaining any state in the central system.¹⁶⁷

Chapman e Agre demonstraram até o momento que a teoria da atividade fundamentada em rotinas é capaz de explicar o desempenho de agentes no ambiente do jogo sem recurso a RM. Entretanto, poder-se-ia dizer o mesmo sobre a cognição? Segundo os autores, sim – pois a cognição em Pengi é simplesmente o registro dos aspectos ativados no jogo pelas unidades lógicas via rotinas. A cada nova interação circunstancial entre sistema periférico e PVR registra-se um novo aspecto que, quando ativado, resultará numa determinada ação na rede combinacional. Resumidamente, a

¹⁶⁴ « Um circuito combinacional é constituído por um conjunto de portas lógicas as quais determinam os valores das saídas diretamente a partir dos valores atuais das entradas. Pode-se dizer que um circuito combinacional realiza uma operação de processamento de informação a qual pode ser especificada por meio de um conjunto de equações Booleanas. No caso, cada combinação de valores de entrada pode ser vista como uma informação diferente e cada conjunto de valores de saída representa o resultado da operação. » José Luís Güntzel e Francisco Assis do Nascimento, “Circuitos Combinacionais”, acessado 24 de maio de 2019, <http://www.inf.ufsc.br/~j.guntzel/isd/isd.html>.

¹⁶⁵ Cfr. seção 1.3 Tipo 3 – a representação mental como ação incorporada (enativismo), 39.

¹⁶⁶ Cfr. seção 1.2 Tipo 2 – a representação mental subsimbólica (conexionismo), 29.

¹⁶⁷ Agre e Chapman, “Pengi”, 1987, 270.

dinâmica cognitiva poderia ser explicada da seguinte forma: dada uma nova interação, imagem e aspecto são respetivamente copiados e registrados no PVR. A seguir, uma resposta é gerada e enviada para a rede – em termos técnicos, uma combinação binária do aspeto processado no PVR é enviado e ficará disponível para a performance da rede combinacional, ou dito de outra forma, na próxima interação a partir da mesma circunstância, o aspecto já estará à disposição da rede.

If the player discovers in a particular situation that *the-bee-on-the-other-side-of-this-block-next-to-me-is-dangerous* because it can easily kick the block into the penguin, this discovery will apply automatically to other specific bees later on that can be described as *the-bee-on-the-other-side-of-this-block-next-to-me*.¹⁶⁸

A cognição é assim o registro e a ativação dos aspetos orquestrados via rotinas visuais. Contudo, é importante destacar que as funções que executam a dinâmica do jogo e que são responsáveis pela execução estão implementadas no circuito combinacional, sendo constituídas de centenas de portas lógicas incrustadas no hardware em Pengi.¹⁶⁹

We are wiring Pengi's central system by hand. Evolution, similarly, wired the central system of insects. But humans and intelligent programs must be able to extend their own networks based on experience with new sorts of situations.¹⁷⁰

3.2.1 A RM do tipo[Y] em Pengi

Conforme foi possível averiguar na seção anterior, a RM não se fez necessária para a agência em Pengi. Poder-se-ia assim supor que o objetivo número 1 desta tese – que os autores antirrepresentacionistas advogam de fato programas afinal representacionistas – sofre de fato uma forte objeção¹⁷¹. Contudo, um olhar mais criterioso na arquitetura da atividade do jogo, mostrará que este modelo também está

¹⁶⁸ Agre e Chapman, 270.

¹⁶⁹ « *Currently, Pengi has a network of several hundred gates and a VRP with about thirty operators.* »; VRP – acrônimo do termo inglês *Virtual Routines Processor*, traduzido nesta tese para PVR, processador virtual de rotinas.

¹⁷⁰ Agre e Chapman, “Pengi”, 1987, 272.

¹⁷¹ Cfr. seção Metodologia, objetivos e estrutura, 20.

fundamentado em RM, conforme a Figura 11 – Modelo da teoria da atividade em Pengi, logo abaixo.

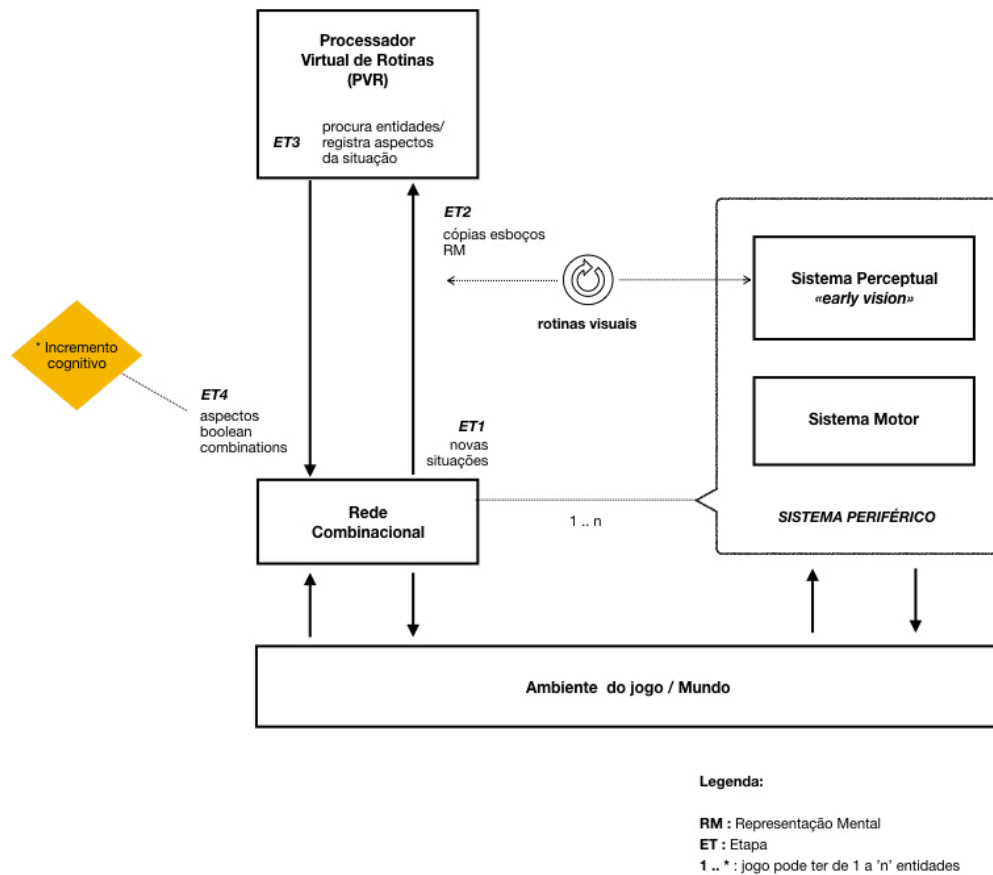


Figura 11 – Modelo da teoria da atividade em Pengi

A partir desta figura, nota-se que o componente PVR tem como uma de suas atribuições, o armazenamento das cópias das circunstâncias vivenciadas pelos sistemas periféricos (ET2). Como resultado deste processo (ET3), o PVR envia as aspectos que serão tratados pela rede combinacional (ET4).

Por conseguinte, segundo a noção de RM definida nesta tese, tem-se o seguinte:

- I. RM possuem intencionalidade, são tipicamente direcionadas para objetos;
- II. RM são encenações do mundo na mente de um agente;
- III. RM fazem parte de estados cognitivos internos.¹⁷²

À vista disto, a tese aponta duas objeções à posição antirrepresentacionista em

¹⁷² Cfr. seção 1.0 Representação mental «RM», 23.

Pengi:

1. As utilizações imediatas dos aspectos pela rede combinacional para as situações já ativadas pela agência caracterizam-se como estados cognitivos internos do sistema (III). Logo, os aspectos são RM.¹⁷³
2. As cópias das circunstâncias mantidas pelo componente PVR caracterizam-se como representações, encenações e reproduções (II) das situações vividas pelos sistemas periféricos. Portanto, as cópias das circunstâncias, ou esboços 2D são RM.

Em relação à objeção 1, os aspetos já tratados estarão automaticamente disponíveis na rede combinacional (ET4) para os sistemas periféricos. Contudo, este automatismo é sempre precedido por uma interação (ET2) e respectivo processamento (ET3) de modo a estar disponível para a agência. Dito de outra forma, o imediatismo da ação somente é possível porque houve um processamento anterior (ET3) da situação, no qual o aspecto, já processado e acessível na rede (ET4), é o estado cognitivo interno do sistema (III).

Com relação a objeção 2, o fato do componente PVR armazenar cópias das circunstâncias vividas pelos agentes (ET2), já constitui um forte argumento a favor do uso de RM, conforme (II). Afinal, uma cópia de uma situação na qual o pinguim foge de uma abelha, nada mais é do que uma representação de um fato que já ocorreu. Por isso, operar sobre a cópia é ainda representar e encenar uma realidade externa (ET2) no PVR. Desta forma, pode-se dizer que a atividade em Pengi é síncrona, o que significa que existe um fluxo de operações necessárias antes da atuação das unidades lógicas. É impossível ao pinguim fugir da abelha, antes do registro do aspecto e da operação sobre a cópia da circunstância – que é uma imagem – referente ao aspecto.

A ação é, em geral, sempre resultado do processamento prévio das circunstâncias. Desse modo, Chapman e Agre constroem uma teoria que realmente contorna a complexidade de representação de modelos e planos na mente dos agentes –

¹⁷³ Esta objeção poderia levantar a hipótese de a emergência no connexionismo também ser caracterizado como estado cognitivo interno da mente. Contudo, no Pengi é explícito que o estado cognitivo é uma combinação booleana, ou seja, sabe-se a natureza do estado. O mesmo não se pode dizer da emergência no cognitivismo. Logo, que a emergência é um estado cognitivo parece não haver dúvida, não se pode dizer o mesmo sobre sua natureza.

Planejamento «P» –, permitindo que a atuação possa ser realizada de forma imediata pela agência – planejamento «p». Contudo, este imediatismo somente é possível após o processamento das circunstâncias, ou em outras palavras, de RM referentes aos aspetos. À vista disto, a questão que se apresenta sobre a cognição neste modelo é a seguinte: O que tem mais relevância na cognição em Pengi – o registrar e ativar os aspectos das circunstâncias ou o ser capaz de processá-las? Não teriam os autores descoberto que o ponto nevrálgico da cognição não é possuir modelos complexos de mundo ou atuar via aspetos, mas sim processar as circunstâncias enquanto RM?

Posto isto, nota-se também uma semelhança entre a estrutura cognitiva presente em Pengi com aquela encontrada para a solução do experimento do bastão: o conexionismo implementacional¹⁷⁴. Voltando ao contexto do experimento, Hayy possuía RM das experiências amargas vividas pela disputa das frutas com os outros animais da ilha. Desprovido de garras, chifres e outras características que lhe permitissem conquistar o almejado alimento, resolve, depois de um certo tempo de reflexão, construir um bastão – o que lhe permitiu contornar o problema.

Este exemplo quando comparado a um evento rotineiro em Pengi, como acompanhar o trajeto de um bloco chutado¹⁷⁵, revela que ambas as dinâmicas são semelhantes. Pois, para acompanhar o trajeto do bloco que o pinguim acabou de chutar, impressa no aspecto *the-block-that-the-block-I-just-kicked-will-collide-with*¹⁷⁶, será necessário que o componente PVR processe através da cópia da imagem da situação vivenciada, ou seja, da RM, os pontos entre o bloco chutado e o seu trajeto até o próximo ponto fixo, para que então a saída – o resultado do processamento – seja tratado na rede combinacional.

¹⁷⁴ Cfr. 2.2 O experimento do bastão, 45.

¹⁷⁵ Cfr. Figura 12 – PVR processando o aspecto *the-block-that-the-block-I-just-kicked-will-collide-with*, 84.

¹⁷⁶ Agre e Chapman, “Pengi”, 1987, 271.

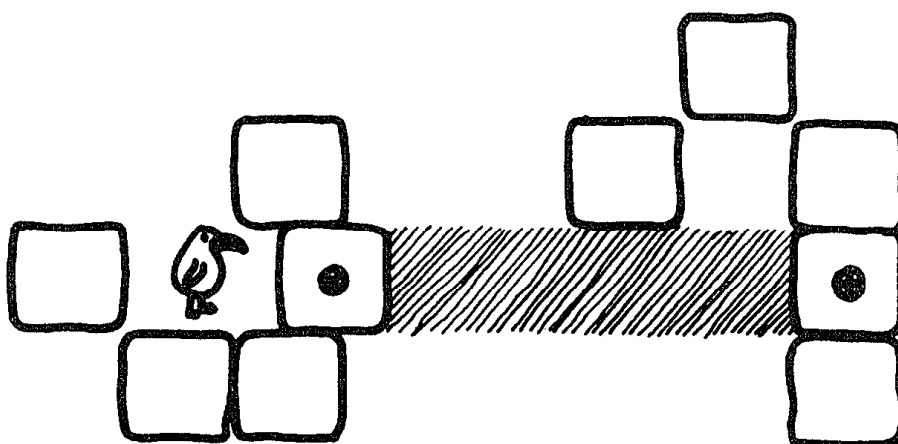


Figura 12 – PVR processando o aspecto *the-block-that-the-block-I-just-kicked-will-collide-with*

Deste modo, tanto planejamentos «P» – construção do bastão defensivo – quanto planejamentos «p» – chutar blocos de gelo – dependerão do processamento de RM para serem executados pelos agentes.

Assim, com base na teoria da atividade em Pengi, a tese definirá como estrutura para as RM do tipo [5], ou indexical implementacional¹⁷⁷, as seguintes ideias:

- I. Os agentes atuam com base em aspectos de circunstâncias e não possuem procedimentos, planos ou teorias;
- II. Cognição é o registro e a ativação dos aspectos das circunstâncias provenientes do ambiente pelo sistema;
- III. Os aspectos são registrados no processador de rotinas visuais (PRV), através de rotinas visuais operantes entre os sistemas periféricos e PRV;
- IV. As atividades são exercidas de forma distribuída, parte pelo sistema central e parte pelos sistemas periféricos;
- V. Os aspectos são representações indexicais e funcionais, pois dependem das circunstâncias vividas e dos propósitos da agência, respectivamente.

¹⁷⁷ Esta tese sempre usará o termo «implementacional» para designar as implementações particulares das RM do tipo [Y] propostas pelos autores antirrepresentacionistas. Logo, uma RM do tipo [Y] implementacional é uma variante particular dos principais programas cognitivistas que apresentam RM em sua estrutura. Neste caso específico, como a estrutura em Pengi apresenta muitos dos princípios indexicais, este tipo de RM será designado como RM indexical implementacional.

- VI. Por apresentar processamento de imagem e manter estados cognitivos no sistema, o tipo indexical implementacional se caracteriza como um tipo de RM.

Por fim, definida as condições de RM do tipo [5] ou indexical implementacional, resta verificar se este tipo atende as condições do experimento do bastão.

3.2.2 Pengi e o experimento do bastão

Com relação a condicionante I da RM do tipo [5], esta parece atender aos requisitos do experimento, pois o fato de um sistema de IA não dispor de procedimentos, planos, ou teorias previamente implantadas acerca de como construir um bastão, coloca-o na mesma posição cognitiva na qual estava Hayy, antes da construção do artefacto. Com relação as rotinas visuais, estas também estão presentes no protagonista – sistema visual – e não seriam um impedimento no experimento.

Na sequência, a condicionante II diz que a cognição é o registro e a ativação dos aspetos das circunstâncias provenientes do ambiente pelo sistema. Isto posto, e sob as lentes da teoria da atividade em Pengi, definir-se-ia a dinâmica do experimento do bastão da seguinte forma:

- i. Hayy precisaria vivenciar uma situação nova (ET1);
- ii. esta circunstância produziria um esboço 2D (ET2);
- iii. que seria enviado, armazenado e operacionalizada no PVR (ET3);
- iv. registrando assim, o aspecto *I'm-fighting-by-fruits* (ET3);
- v. no final deste processo, este aspeto seria tratado no ambiente, e neste exemplo, dar-se-ia a ação de construir um bastão (ET4).¹⁷⁸

Contudo, conforme exposto na seção anterior, o processamento do esboço, ou da imagem da circunstância tem uma importância chave no processo cognitivo da RM do tipo [5]. Tecnicamente, seguindo as diretrizes técnicas do que foi implementado em Pengi, mostrou-se que o processador de rotinas visuais é constituído de operadores lógicos¹⁷⁹. Ou seja, a forma como os operadores estão estruturados na rede

¹⁷⁸ Cfr. Figura 11 – Modelo da teoria da atividade em Pengi, 81.

¹⁷⁹ Agre e Chapman, “Pengi”, 1987, 272.

combinacional é o que permite ao PVR encontrar as entidades e registrar os aspectos das circunstâncias. Todavia, de acordo com as condições definidas no experimento do bastão, o sistema de IA não poderia possuir, de antemão, das funções ou da estrutura lógica preexistente para processar a produção do bastão. Este incremento deveria ocorrer *a posteriori* e de forma autônoma – sem a interferência externa e técnica de qualquer tipo.

Além disto, a própria performance do aspecto *I'm-fighting-by-fruits* dependeria da atualização da rede combinacional, composta por centenas de portas lógicas¹⁸⁰ – responsáveis pelo desempenho do jogo. Portanto, o PVR e a rede combinacional são capazes de interpretar e atuar os aspectos a partir das diretrizes definidas pelos engenheiros técnicos. Assim, se por acaso fosse colocado aleatoriamente no labirinto do jogo, um coqueiro, o processador não saberia como interpretá-lo, a menos que novas portas lógicas fossem acrescentadas ou alteradas no PVR. Por conseguinte, a mesma explicação seria válida para a performance da rede combinacional, pois não seria possível as unidades lógicas – pinguins e abelhas – arremessarem cocos uns contra os outros sem a atualização da rede combinacional.

Da mesma forma que a inteligência das criaturas de Brooks é determinada pela estrutura das máquinas finitas, o mesmo acontece em Pengi, que apesar de possuir sistema cognitivo, tal está limitado ao número determinado e finito de saídas, resultantes da forma de organização das portas lógicas implantadas no processador visual de rotinas. Como contornar a rigidez imposta pela determinação eletrônica/física? No livro *Infinite powers – How Calculus Reveals the Secrets of the Universe*¹⁸¹, o autor Steven Strogatz defende a tese de que o cálculo infinito, ou mais precisamente, a redução infinita de um problema é a chave para a resolução de todas as questões do universo. O autor dá o exemplo da complexidade do cálculo de um círculo, que sob as lentes da disciplina do cálculo, não se apresenta através de uma forma redonda perfeita, mas sim pela composição infinitesimal de pequenas retas. Segundo o autor, esta seria a fórmula para o cálculo universal, dividir um problema *ad infinitum*.

¹⁸⁰ Agre e Chapman, 272.

¹⁸¹ Steven Strogatz, *Infinite Powers: How Calculus Reveals the Secrets of the Universe* (New York: Houghton Mifflin Harcourt, 2019).

It's an obvious point that a circle is round, it doesn't contain any straight segments. But in calculus we pretend that it does. We pretend that a circle, if you zoomed in on it under a microscope, you'd see that the circle would start to look straighter and straighter under increasing magnification. It's almost as if it were made up of many infinitesimally small straight pieces.

This is the great brainstorm of calculus. Think about an object that's complicated and pretend it's made up of much simpler pieces, at the cost of having to imagine infinitely many of them. So harnessing infinity turned out to be the key to solving lots of problems that were previously intractable.¹⁸²

Supondo que esta seja uma das características responsáveis pelo sucesso da autonomia cognitiva, como seria possível reproduzi-la em sistemas de IA? Ao que tudo indica, o processamento indeterminado de RM, ou o cálculo infinito conforme aponta Steven Strogatz, apresenta-se como forte evidência corroborativa da autonomia cognitiva e que a RM indexical implementacional e a RM enativista implementacional não conseguiram endereçar. Portanto, para Pengi apresentar autonomia cognitiva seria necessário que o PVR fosse capaz de realizar cálculos infinitesimais com as infinitas circunstâncias possíveis de serem capturadas do ambiente real.

Pelas razões apontadas acima, o tipo de RM [5], ou indexical implementacional, também não preenche aos requisitos definidos para o experimento do bastão e, portanto, não permite a autonomia cognitiva em sistemas de IA.

¹⁸² Steven Strogatz, No, really, calculus can be beautiful and this mathematician will tell you why | CBC Radio, entrevista de Bob McDonald, 24 de maio de 2019, <https://www.cbc.ca/radio/quirks/may-25-sharks-on-a-bird-diet-fossils-of-fungus-lifelike-machines-and-more-1.5147055/no-really-calculus-can-be-beautiful-and-this-mathematician-will-tell-you-why-1.5147060>.

Capítulo 4 – Conclusão

O objetivo principal da tese foi averiguar qual o tipo de RM seria necessário para a autonomia cognitiva em sistemas de IA. A partir deste tema, foi elaborada a seguinte questão:

Que relação há entre sistemas de IA possuírem estrutura[s] de RM do[s] tipo[s] [X] e tais sistemas apresentarem autonomia cognitiva?

Esta questão permitiu identificar duas proposições básicas:

p: Sistemas de IA apresentam autonomia cognitiva

q: Sistemas de IA possuem estrutura[s] de RM do[s] tipo[s] [X]

Estas duas proposições, por sua vez, podem estar relacionadas dos seguintes dois modos:

P1 – Possuir estrutura de RM do tipo [X] é condição necessária para a autonomia cognitiva em sistemas de IA. ($p \rightarrow q$)

P2 – Sistemas de IA têm autonomia cognitiva, mas não possuem estrutura[s] de RM do[s] tipo[s] [X]. ($p \wedge \sim q$)

Conforme exposto na introdução desta dissertação, a forma mais eficiente de encontrar a RM do tipo [X] foi averiguando aqueles programas que fazem objeção a necessidade da RM para a cognição em sistemas de IA. Consequentemente, tais programas deveriam oferecer propostas capazes de explicar a cognição sem o artifício da RM. Além disso, esta tese assumiu a posição de que os programas que não dispunham de RM no processo cognitivo, ou seja, os programas antirrepresentacionistas, utilizavam-se inevitavelmente de algum tipo de RM, identificados provisoriamente como sendo de tipo [Y]. Isto posto, restou averiguar se a[s] RM[s] do[s] tipo[s] [Y] atenderiam a necessidade da autonomia cognitiva em sistemas de IA, e se tais tipos se assemelhavam de alguma forma como o tipo de RM identificado em Hayy.

Isto posto, caso a «P2» fosse falsa, os seguintes objetivos seriam apurados:

Objetivo 1: A tese irá mostrar que os programas antirrepresentacionistas usam RM do[s] tipo[s] [Y] e, portanto, também são representacionistas, diferenciados

apenas pelos tipos de RM.

Objetivo 2: A tese irá investigar se as RM do[s] tipo[s] [Y] identificadas nos textos, satisfazem a necessidade de autonomia cognitiva em sistema de IA, através do experimento elaborado a partir de um excerto do livro de Tufayl, denominado «o experimento do bastão».¹⁸³

Objetivo 3: A tese irá apontar através dos resultados obtidos nos objetivos 1 e 2, quais RM do[s] tipo[s] [Y] são necessárias para a autonomia cognitiva em sistema de IA, e desta forma, classificá-las como RM do[s] tipo[s] [X].

Objetivo 4: A tese irá apontar através dos resultados obtidos nos objetivos 1 e 2, quais RM do[s] tipo[s] [Y] não são necessárias para a autonomia cognitiva em sistemas de IA e desta forma, excluí-las como RM do[s] tipo[s] [X].

Restabelecida as questões e objetivos da tese, este último capítulo terá a incumbência de apresentar a apreciação final de cada um desses itens, de acordo com a averiguação realizada no decurso dos capítulos predecessores.

4.1 Apreciação da proposição «P2» e dos seus objetivos

A proposição «P2» se mostrou falsa, devido a incapacidade das teorias antirrepresentacionistas – RM do tipo [4], ou enativismo implementacional e RM do tipo [5], ou indexical implementacional – satisfazerem as condições impostas no experimento do bastão. Por «P2» ser falsa, os objetivos abaixo puderam ser considerados e apurados.

Objetivo 1: A tese irá mostrar que os programas antirrepresentacionistas usam RM do[s] tipo[s] [Y] e, portanto, também são representacionistas, diferenciados apenas pelos tipos de RM.

Conclusão: Conforme foi possível averiguar na seção 3.1.1 A RM do tipo[Y] das criaturas de Brooks, o autor tutelou que a arquitetura de subsunção permitiria contornar a necessidade de RM na cognição e na inteligência dos robôs, permitindo-lhes atuarem no ambiente sem instâncias de programas, regras condicionais ou quaisquer teorias acopladas a eles. Contudo, devido ao fato de a estrutura dos robôs ser constituída de

¹⁸³ Cfr. seção 2.2 O experimento do bastão,45.

máquinas de estados finitas (MEF), e tais modelos serem abstrações conceituais de interação, permitiram a inferência de que a RM estaria marcada na própria estrutura física do agente, ou em outras palavras, a RM neste tipo está incrustada no hardware ao invés de no software. Consequentemente, estas razões somadas à importância da estrutura corporal na inteligência e na cognição na teoria de Brooks, definiram a RM deste programa como tipo [4] ou enativismo implementacional.

Já na seção 3.2.1 A RM do tipo[Y] em Pengi, vimos os autores apresentaram uma implementação teórica fundamentada em rotinas ao invés de procedimentos, dispensando a necessidade da RM tanto na interação com o ambiente quanto na cognição. Todavia, após averiguar a arquitetura foi possível constatar que o processador visual de rotinas (PVR) e a rede combinacional (sistemas periféricos) apresentavam características similares àquelas definidas nas RM¹⁸⁴, conforme as objeções:

1. As utilizações imediatas dos aspectos pela rede combinacional para as situações já vivenciadas pelas unidades lógicas caracterizam-se como estados cognitivos internos do sistema (III). Logo, os aspectos são RM.
2. As cópias das circunstâncias mantidas pelo componente PVR caracterizam-se como representações, encenações e reproduções (II) das situações vividas pelos sistemas periféricos. Portanto, as cópias das circunstâncias, ou esboços 2D são RM.

Devido ao caráter «indexical» dos aspectos – dependem da circunstância para fazerem sentido –, este tipo estrutural de RM foi definido como tipo [5], ou indexical implementacional.

Por fim, as RM do tipo[s] [Y] identificadas nos autores antirrepresentacionistas foram: RM do tipo [4] ou enativismo implementacional e RM do tipo [5], ou indexical implementacional.

¹⁸⁴ Cfr. as definições da seção 1.0 Representação mental «RM», 23.

Objetivo 2: A tese irá investigar se as RM do[s] tipo[s] [Y] identificadas nos textos, satisfazem a necessidade de autonomia cognitiva em sistema de IA, através do experimento elaborado a partir de um excerto do livro de Tufayl, denominado «o experimento do bastão».¹⁸⁵

Conclusão: As RM dos tipos [Y] identificadas no Objetivo 1, não atenderam as condições para a autonomia cognitiva em sistemas de IA definidas no experimento do bastão. Assim, foi possível averiguar na seção 3.1.2 As criaturas de Brooks e o experimento do bastão, que a autonomia cognitiva somente seria possível caso ocorresse o incremento da habilidade «construir bastão» na arquitetura de subsunção. Porém, isto violaria a condicionante de «autonomia» definida no experimento, que impõe restrição de atualização e interferência externa no processo cognitivo¹⁸⁶. Por esse motivo, a arquitetura de subsunção não satisfaz as necessidades do experimento.

Por conseguinte, na seção 3.2.2 Pengi e o experimento do bastão foi possível averiguar que o modelo intelectual e cognitivo baseado em rotinas também não satisfaz as condicionantes do experimento, pois seria necessário atualizar o processador de rotinas visuais (PVR) para se atingir a proeza cognitiva de Hayy. Como a cognição é determinada pelo registro dos aspectos pelo PVR, não seria possível processar circunstâncias distintas àquelas já implementada no processador, a não ser que ocorresse uma atualização do PVR.

Portanto, as RM do tipo[s] [Y] – tipo [4] ou enativismo implementacional e tipo [5], ou indexical implementacional –, não satisfizeram as necessidades da autonomia cognitiva em sistemas de IA.

Objetivo 3: A tese irá apontar através dos resultados obtidos nos objetivos 1 e 2, quais RM do[s] tipo[s] [Y] são necessárias para a autonomia cognitiva em sistemas de IA, e desta forma, classificá-las como RM do[s] tipo[s] [X].

Conclusão: Conforme os resultados dos objetivos 1 e 2, nenhuma das RM dos tipos [Y] – RM do tipo [4] ou enativismo implementacional e RM do tipo [5] ou indexical implementacional –, atenderam a autonomia cognitiva em sistemas de IA e

¹⁸⁵ Cfr. seção 2.2 O experimento do bastão, 46.

¹⁸⁶ Cfr. seção Observações sobre palavras-chave, 16.

puderam ser classificados como RM do[s] tipo[s] [X].

Objetivo 4: A tese irá apontar através dos resultados obtidos nos objetivos 1 e 2, quais RM do[s] tipo[s] [Y] não são necessárias para a autonomia cognitiva em sistemas de IA e desta forma, excluí-las como RM do[s] tipo[s] [X].

Conclusão: Conforme os resultados dos objetivos 1 e 2, as RM do tipo [4] ou enativismo implementacional e RM do tipo [5] ou indexical implementacional, não são necessárias para a autonomia cognitiva em sistemas de IA e portanto, foram excluídas como RM do[s] tipo[s] [X].

4.2 Apreciação da proposição «P1»

Devido ao princípio da contradição, a proposição «P1» foi considerada verdadeira já que a proposição «P2» foi apurada como falsa:

P1– Possuir estrutura de RM do tipo [X] é condição necessária para a autonomia cognitiva em sistemas de IA. ($p \rightarrow q$)

O valor da «P1» é corroborada pela RM do tipo conexionista implementacional, o único tipo, dentre aqueles averiguados, capaz de satisfazer as condicionantes do experimento do bastão.

Portanto, esta tese assume que o tipo [X] de RM responsável pela autonomia cognitiva dos agentes é a RM do tipo conexionista implementacional. Por fim, os sistemas de IA somente serão capazes dessa proeza cognitiva se utilizarem-se de equivalente tipo de RM, ou dito em outras palavras, da posse e processamento indeterminado de RM.

4.3 Semelhanças entre os Tipos de RM [Y] e a RM em Hayy

Há semelhanças entre o tipo de RM utilizado por Hayy, conexionismo implementacional, e a RM do tipo [5] ou indexical implementacional, identificada em Pengi. No experimento do bastão, o sucesso da autonomia cognitiva em Hayy foi possível devida a posse e o processamento «indeterminado» de RM. Funcionamento análogo foi encontrado na arquitetura cognitiva em Pengi, cujo processador visual de rotinas (PVR) porta os esboços produzidos pelos sistemas periféricos e processa-os de

forma «determinada», para então disponibilizar os aspectos que serão atuados pela rede combinacional.

O que ficou evidente é que o sucesso da autonomia cognitiva, a partir dos autores estudados nesta tese, está relacionada a dois fatores:

1. Posse de RM;
2. Processamento indeterminado de RM.

Desse modo, torna-se necessário investigar quais são as características funcionais e físicas do processamento de RM utilizada pela estrutura cognitiva humana, para que então se possa avaliar a possibilidade de o mesmo tipo de processo ser reproduzido em sistemas de IA. O que se pode supor previamente é que a velocidade de processamento e as características físicas não são suficientes para a autonomia cognitiva. Outras características, como por exemplo, a redução de um problema *ad infinitum*¹⁸⁷, indicam possíveis caminhos para a continuidade da investigação.

O próximo passo, portanto, seria averiguar: as características de processamento informacional necessárias para a autonomia cognitiva em sistemas de IA.

4.4 Implicações em outras áreas dos resultados apurados

Esta tese teve como objetivo investigar e responder a seguinte pergunta: Que relação há entre sistemas de IA possuírem estrutura[s] de RM do[s] tipo[s] [X] e tais sistemas apresentarem autonomia cognitiva? Por conseguinte, esta pergunta leva a uma outra não menos importante: Quais são as consequências de um sistema de IA utilizar o mesmo tipo de RM que a agência humana? Estas questões serão refletidas dentro de duas abordagens contemporâneas ligadas ao tema: a singularidade e a *open letter on AI*¹⁸⁸.

A singularidade é a teoria que tutela que a tecnologia está avançando de forma tão rápida e intensa, que suas implicações inevitavelmente irão afetar todas as dimensões da

¹⁸⁷ Strogatz, *Infinite Powers: How Calculus Reveals the Secrets of the Universe*.

¹⁸⁸ “AI Open Letter”, Future of Life Institute, acessado 27 de maio de 2019, <https://futureoflife.org/ai-open-letter/>.

humanidade, inclusive a biológica¹⁸⁹. Desse modo, num futuro próximo, a singularidade irá permitir que as limitações do corpo biológicos sejam transpostas, bem como a erradicação dos problemas sociais, econômicos e ambientais. Segundo os seus defensores, até a barreira da morte poderá ser transposta quando houver a fusão entre biologia e tecnologia. «*Our mortality will be in our own hands.*»¹⁹⁰

Segundo o futurologista, inventor e diretor do projeto *Google Brain*¹⁹¹, Ray Kurzweil, nossa inteligência irá conseguir transpor a limitada capacidade de processamento informacional, que demanda de trilhões de neurônios para conseguir identificar um determinado padrão. Com o advento da singularidade, esta capacidade será expandida para trilhões de trilhões de vezes da capacidade atual da mente, proporcionando a humanidade um poder de processamento de informação que mudará todos os patamares conhecidos atualmente. No ápice da singularidade, a espécie humana será fundida junto à tecnologia, ocorrendo então uma transposição biológica que lançará este novo ser a conquista do universo.

Although impressive in many respects, the brain suffers from severe limitations. We use its massive parallelism (one hundred trillion interneuronal connections operating simultaneously) to quickly recognize subtle patterns. But our thinking is extremely slow: the basic neural transactions are several million times slower than contemporary electronic circuits. [...] The Singularity will allow us to transcend these limitations of our biological bodies and brains. [...] By the end of this century, the nonbiological portion of our intelligence will be trillions of trillions of times more powerful than unaided human intelligence.¹⁹²

Partindo de uma posição menos futurista e mais pragmática, a *open letter on AI*¹⁹³ é um documento assinado por mais de 8000 signatários, dentro os quais, Elon Musk e Stephen Hawkins. O objetivo é definir as prioridades de pesquisas relacionadas a IA,

¹⁸⁹ « *Although the Singularity has many faces, its most important implication is this: our technology will match and then vastly exceed the refinement and suppleness of what we regard as the best of human traits.* » Walter Russell Mead e Ray Kurzweil, “The Singularity Is near: When Humans Transcend Biology”, *Foreign Affairs* 85, nº 3 (2006): 24, <https://doi.org/10.2307/20031996>.

¹⁹⁰ Mead e Kurzweil, 24.

¹⁹¹ Foer, *Mundo sem mente: a ameaça existencial da alta tecnologia*, 65.

¹⁹² Mead e Kurzweil, “The Singularity Is Near”, 24.

¹⁹³ “AI Open Letter”.

assim como alertar e debater com a sociedade civil e acadêmica sobre as implicações da expansão exponencial dos sistemas de IA em todas as dimensões humanas: economia, ciências, educação, pesquisa espacial, somente para citar algumas áreas, além do impacto ambiental dessa expansão.

Voltando ao tema da tese, a estrutura de RM adequada para a autonomia cognitiva parece ser o ponto nevrálgico de todo este debate. Caso seja possível reproduzir a estrutura de RM correta para a autonomia cognitiva em sistemas de IA, as reivindicações da singularidade e da *open letter on AI* são corroboradas com evidências científicas de que seus anseios, preocupações e previsões provavelmente irão acontecer. Entretanto, caso não seja possível replicar este tipo de estrutura de RM em sistema de IA, ambas teorias precisariam ser revisadas e no caso da singularidade, talvez até mesmo descartada, uma vez que não seria possível a IA superar a racionalidade humana.

Além disto, vale ressaltar que há entre a comunidade acadêmica uma dissonância sobre os limites cognitivos em sistemas de IA. Por um lado, Ray Kurzweil coloca que a capacidade de processamento informacional, como uma das condições para que a mente humana seja ultrapassada pela IA; assim como Brooks apostou na autonomia dos sistemas perceptuais e motores de suas criaturas e Chapman e Agre na interação e cognição baseada em rotinas. Em contrapartida, pesquisadores como Manuela Veloso¹⁹⁴ e Sangbae Kim¹⁹⁵ apontam que a próxima fase da IA será de simbiose e cooperação, devido ao entendimento de que os sistemas de IA afinal tem os seus limites.

Em vista disto, uma obra do século XII, *O filósofo autodidata*¹⁹⁶, permite que no século XXI sejam colados à prova todas estas teorias, para então se chegar à conclusão de que os sistemas de IA estão muito longe da capacidade cognitiva de Hayy, capaz de sobreviver e evoluir cognitivamente de forma autônoma e sozinho em uma ilha deserta. A que conclusões finais se pode chegar? Pode-se afirmar que a cognição muito provavelmente não seja uma questão de velocidade de processamento, como sugere Kurzweil, de locomoção e percepção do ambiente como apontou Brooks e nem de

¹⁹⁴ Ver histórico profissional de Manuela M. Veloso na nota 131, 68.

¹⁹⁵ “MECHE PEOPLE: Sangbae Kim | MIT Department of Mechanical Engineering”.

¹⁹⁶ Tufayl, *O Filósofo Autodidata*.

interação via rotinas como foi apresentando em Pengi.

O sucesso para a autonomia cognitiva é possuir a estrutura adequada de RM. Resta à ciência e a filosofia conhecer a sua natureza e funcionamento, para quiçá tentar reproduzi-la em sistemas de IA. Caso não seja possível, só restará ao homem contentar-se com a ideia de que a transposição da espécie *homo sapiens* pela *homo technologicus*, seja afinal, inviável, e conseqüentemente, enfrentar as implicações disto.

Bibliografia

Fontes primárias

Philip E. Agre e David Chapman, “Pengi: An Implementation of a Theory of Activity”, in *Proceedings of the Sixth National Conference on Artificial Intelligence - Volume 1*, AAAI’87 (AAAI Press, 1987), 268–272, <http://dl.acm.org/citation.cfm?id=1863696.1863744>.

Brooks, Rodney A. “Achieving Artificial Intelligence through Building Robots”, AIM–899, 1º de maio de 1986. <https://dspace.mit.edu/handle/1721.1/6451>.

———. “Intelligence Without Representation”. *Artificial Intelligence* 47, nº 1–3 (janeiro de 1991): 139–59. [https://doi.org/10.1016/0004-3702\(91\)90053-M](https://doi.org/10.1016/0004-3702(91)90053-M).

Foer, Franklin. Mundo sem mente: a ameaça existencial da alta tecnologia. Traduzido por Luís Oliveira Santos. 1o ed. Temas e Debates – Círculo de Leitores, 2018.

Güntzel, José Luís, e Francisco Assis do Nascimento. “Circuitos Combinacionais”. Acessado 24 de maio de 2019. <http://www.inf.ufsc.br/~j.guntzel/isd/isd.html>.

Harari, Yuval Noah. “The Tree of Knowledge”. In *Sapiens: a brief history of humankind*, 36–. United States, 2015.

Harbisson, Neil. *I Listen to Color*. Acessado 15 de abril de 2019. https://www.ted.com/talks/neil_harbisson_i_listen_to_color.

Mead, Walter Russell, e Ray Kurzweil. “The Singularity Is near: When Humans Transcend Biology”. *Foreign Affairs* 85, nº 3 (2006): 160. <https://doi.org/10.2307/20031996>.

Merleau-Ponty, Maurice. “The Experience of the Body and Classical Psychology”. In *Phenomenology of Perception*, 103–11. London: Routledge, 2002.

Neumeyer, C. “Das Farbensehen des Goldfishes”. Ph. D. dissertation, University of Mainz, 1986.

“Premium Conference – Fatores para o Sucesso da Interação com a Inteligência Artificial | Universidade Católica Portuguesa – Porto”. Acessado 12 de fevereiro de 2019. <http://www.porto.ucp.pt/pt/central-eventos/premium-conference-fatores-para-sucesso-interacao-com-inteligencia-artificial>.

Strogatz, Steven. *Infinite Powers: How Calculus Reveals the Secrets of the Universe*. New York: Houghton Mifflin Harcourt, 2019.

———. No, really, calculus can be beautiful and this mathematician will tell you why | CBC Radio. Entrevista de Bob McDonald, 24 de maio de 2019.
<https://www.cbc.ca/radio/quirks/may-25-sharks-on-a-bird-diet-fossils-of-fungus-lifelike-machines-and-more-1.5147055/no-really-calculus-can-be-beautiful-and-this-mathematician-will-tell-you-why-1.5147060>.

Tufayl, Ibn. *O Filósofo Autodidata*. Traduzido por Isabel Loureiro. Edição: 1. São Paulo: UNESP, 2005.

Turing, Alan. “Computing machinery and intelligence”. *Mind* 50 (1950): 433–60.

———. “On computable numbers, with an application to the Entscheidungsproblem”. *Proc. London Maths. Soc.*, 2, 42 (1936): 230–65.

Varela, Francisco J., Evan Thompson, e Eleanor Rosch. “Emergent properties and connectionism”. In *The Embodied Mind: Cognitive Science and Human Experience* 2, Revised edition., 149–69. Cambridge, Massachusetts ; London England: MIT Press, 1991.

———. “Emergent properties and connectionism”. In *The Embodied Mind: Cognitive Science and Human Experience*, Revised edition., 85–103. Cambridge, Massachusetts ; London England: MIT Press, 2016.

———. “Enaction: Embodied Cognition”. In *The Embodied Mind: Cognitive Science and Human Experience*, Revised edition., 147–83. Cambridge, Massachusetts ; London England: MIT Press, 2016.

———. “Evolutionary Path Making and Natural Drift”. In *The Embodied Mind: Cognitive Science and Human Experience*, Revised edition., 185–213. Cambridge, Massachusetts ; London England: MIT Press, 2016.

———. “Symbols: the cognitivist hypothesis”. In *The Embodied Mind: Cognitive Science and Human Experience*, Revised edition., 37–57. Cambridge, Massachusetts ; London England: MIT Press, 2016.

Varela, Dr Drauzio. “Neurônio”. Portal Drauzio Varela, 6 de maio de 2014.
<https://drauzioharela.uol.com.br/corpo-humano/neuronio/>.

Veloso, Manuela. “Fatores para o Sucesso da Interação com a Inteligência Artificial”. Universidade Católica Portuguesa, 2018.

Fontes secundárias

“AI Open Letter”. Future of Life Institute. Acessado 27 de maio de 2019.

<https://futureoflife.org/ai-open-letter/>.

“Cognoscere”. In *Oxford Latin Dictionary*, 346. Oxford: Oxford University Press, 1968.

“Cor e Visão – Carlos Fiolhais”. *Ano Internacional da Luz 2015* (blog), 22 de junho de 2015. <http://ail2015.org/index.php/2015/06/22/cor-e-visao-carlos-fiolhais/>.

“Finite State Machines | Brilliant Math & Science Wiki”. Acessado 2 de maio de 2019.

<https://brilliant.org/wiki/finite-state-machines/>.

Franklin, Stan. *Artificial Minds*. A Bradford Book. Cambridge, Mass.: MIT Press, 2001.

Garson, James. “Connectionism”. Organizado por Edward N Zalta, Uri Nodelman, e Colin Allen. *The Stanford Encyclopedia of Philosophy*, 2018, 20.

Giaimo, Cara. “Sangbae Kim Ditched the Rules of Robotics—and Built Tomorrow’s first Responder.”, [s.d.], 18.

Hao, Karen. “We Analyzed 16,625 Papers to figure out Where AI Is Headed Next”. *MIT Technology Review*, 2019, 14.

iRobot. “Robô Aspirador de Pó Inteligente Roomba 960 iRobot”. Acessado 1º de maio de 2019. <https://www.irobotloja.com.br/roomba-960-robo-aspirador-de-po-inteligente-bivolt-irobot-r960400/p>.

Knight, Will. “The Demise of Rethink Robotics Shows How Hard It Is to Make Machines Truly Smart”. *MIT Technology Review*, 4 de outubro de 2018, 16.

Liesbeth, De Mol. “Turing Machines”. Organizado por Zalta. *The Stanford Encyclopedia of Philosophy*, nº Winter 2018 (2018).

<https://plato.stanford.edu/archives/win2018/entries/turing-machine/>.

“MECHE PEOPLE: Sangbae Kim | MIT Department of Mechanical Engineering”.

Acessado 3 de maio de 2019.

<http://meche.mit.edu/people/faculty/SANGBAE@MIT.EDU>.

“Open Letter on Artificial Intelligence”. In *Wikipedia*, 4 de abril de 2018.

https://en.wikipedia.org/w/index.php?title=Open_Letter_on_Artificial_Intelligence&oldid=834163483.

“Oxana Malaya”. In *Wikipédia, a enciclopédia livre*, 11 de maio de 2019.

https://pt.wikipedia.org/w/index.php?title=Oxana_Malaya&oldid=55110507.

“Pengo (Video Game)”. In *Wikipedia*, 14 de março de 2019.
[https://en.wikipedia.org/w/index.php?title=Pengo_\(video_game\)&oldid=887735053](https://en.wikipedia.org/w/index.php?title=Pengo_(video_game)&oldid=887735053).

Pitt, David. “Mental Representation”. Organizado por Edward N Zalta, Uri Nodelman, e Colin Allen. *The Stanford Encyclopedia of Philosophy*, Winter Edition de 2018, 42.

Rescorla, Michael. “The Computational Theory of Mind”. Organizado por Edward N Zalta, Uri Nodelman, e Colin Allen, 2017, 40.

“Rethink Robotics”. In *Wikipedia*, 5 de janeiro de 2019.
https://en.wikipedia.org/w/index.php?title=Rethink_Robotics&oldid=876974751.
“rodney brooks – Pesquisa Google”. Acessado 11 de fevereiro de 2019.
<https://www.google.com/search?kgmid=/m/02ysh0&hl=pt-PT&pgs=2ea04802300be52f&q=rodney+brooks&shndl=0&source=sh/x/kp&entrypoint=sh/x/kp>.

Russell, Stuart, Daniel Dewey, e Max Tegmark. “Research Priorities for Robust and Beneficial Artificial Intelligence”. *AI Magazine* 36, nº 4 (31 de dezembro de 2015): 105.
<https://doi.org/10.1609/aimag.v36i4.2577>.

“Sawyer Collaborative Robots for Industrial Automation”. Rethink Robotics. Acessado 3 de maio de 2019. <https://www.rethinkrobotics.com/sawyer/>.

Siegel, Susanna. “The Contents of Perception”. In *The Stanford Encyclopedia of Philosophy*, organizado por Edward N. Zalta, Winter 2016. Metaphysics Research Lab, Stanford University, 2016.
<https://plato.stanford.edu/archives/win2016/entries/perception-contents/>.

“Sistemas Computacionais”. Acessado 3 de abril de 2019.
<http://uab.ifsul.edu.br/tsiad/conteudo/modulo1/sop/ua/at2/>.

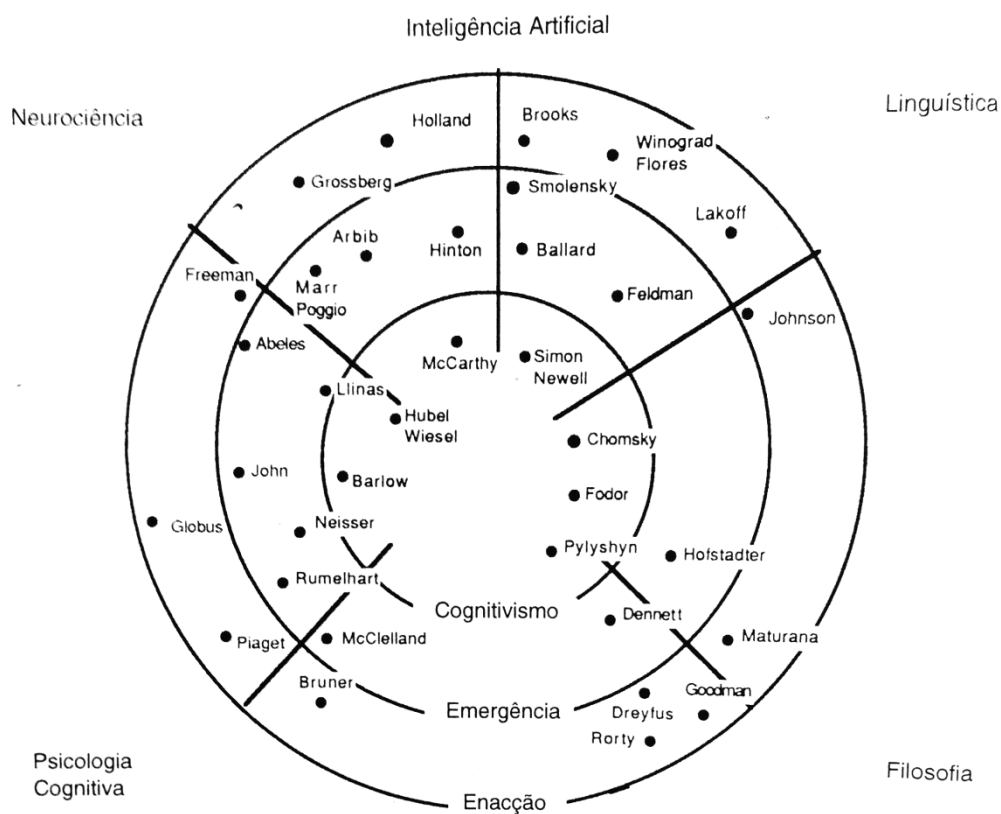
Tufayl, Ibn. *Hayy Ibn Yaqzān*. Editado e traduzido por Leen Evan Goodman. 5º ed. Los Angeles: Gee tee bee, 2003.

“Von Neumann architecture – Wikipedia, the free encyclopedia”. Acessado 21 de março de 2019.
http://jupiter.math.nctu.edu.tw/~weng/courses/IC_2007/PROJECT_MATH_CLASS1/2_HISTORY_ALL_FROM_AN_MATHEMATICAN/Von%20Neumann%20architecture%20-%20Wikipedia,%20the%20free%20encyclopedia.htm.

Wilson, Robert A., e Lucia Foglia. “Embodied Cognition”. Organizado por Edward N Zalta, Uri Nodelman, e Colin Allen. *The Stanford Encyclopedia of Philosophy*, Spring Edition de 2017, 81.

Anexos

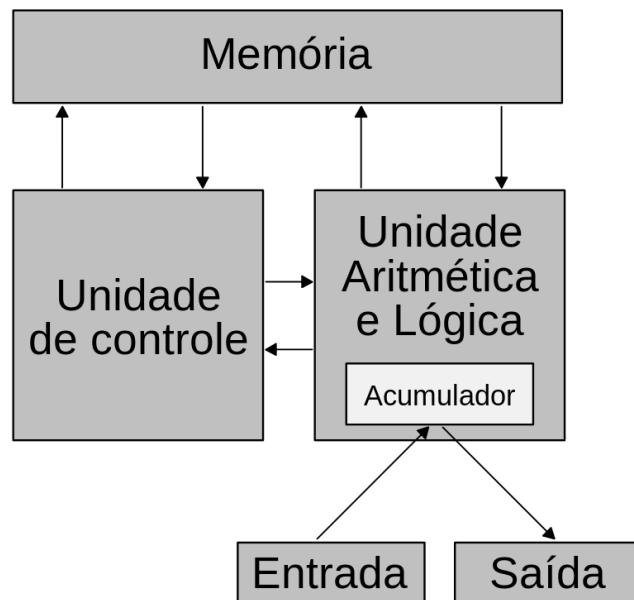
Anexo 1 – Diagrama conceitual das ciências cognitivas



Anexo 1 – Diagrama conceitual das ciências cognitivas¹⁹⁷

¹⁹⁷ Francisco J. Varela, Evan Thompson, e Eleanor Rosch, “Emergent properties and connectionism”, in *The Embodied Mind: Cognitive Science and Human Experience 2*, revised edition (Cambridge, Massachusetts ; London England: MIT Press, 1991), 69.

Anexo 2 – Arquitetura de Von Neumann



Anexo 2 – Arquitetura de von Neumann¹⁹⁸

¹⁹⁸ “Von Neumann architecture – Wikipedia, the free encyclopedia”, acessado 21 de março de 2019, http://jupiter.math.nctu.edu.tw/~weng/courses/IC_2007/PROJECT_MATH_CLASS1/2_HISTORY_ALL_FROM_AN_MATHEMATICAN/Von%20Neumann%20architecture%20-%20Wikipedia,%20the%20free%20encyclopedia.htm.

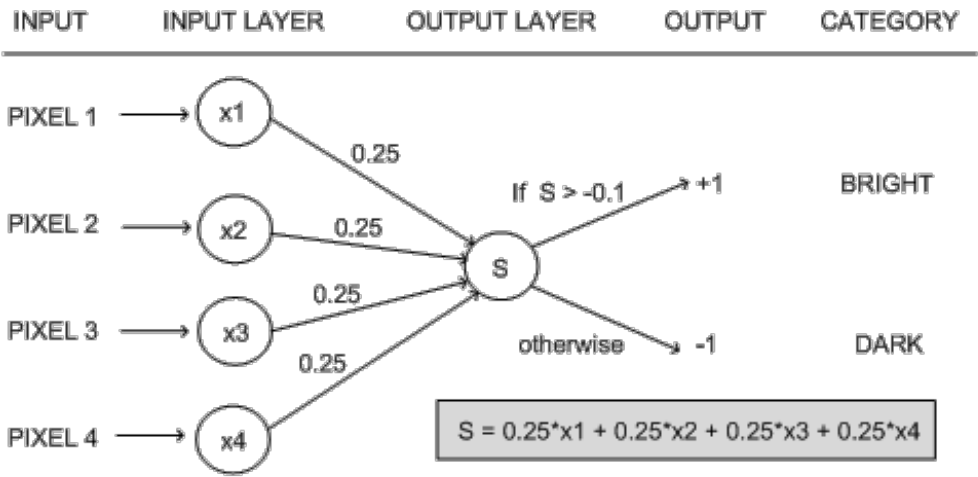
Anexo 3 – Modelo esquemático de um sistema computacional



Anexo 3 – Modelo esquemático de um sistema computacional¹⁹⁹

¹⁹⁹“Sistemas Computacionais”, acessado 3 de abril de 2019, <http://uab.ifsul.edu.br/tsiad/conteudo/modulo1/sop/ua/at2/>.

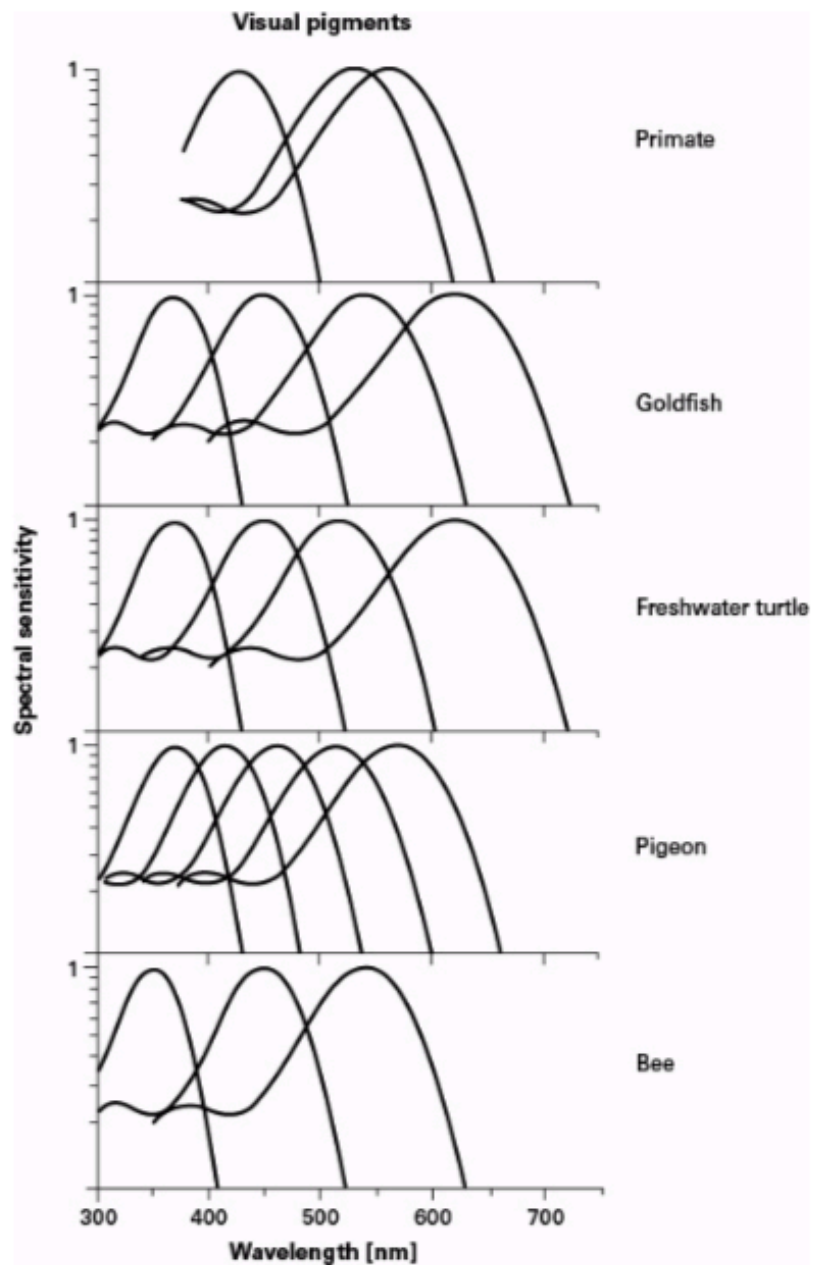
Anexo 4 – Exemplo de classificação de um percetrão



Anexo 4 – Exemplo de classificação de um percetrão²⁰⁰

²⁰⁰ SAGAR SHARMA, “What the Hell is Perceptron?”, Towards Data Science, 9 de setembro de 2017, <https://towardsdatascience.com/what-the-hell-is-perceptron-626217814f53>.

Anexo 5 – Mecanismo tetrocromático versus mecanismo tricromático



Anexo 5 – Mecanismo tetrocromático versus mecanismo tricromático²⁰¹

²⁰¹ Mecanismo tetrocromático versus mecanismo tricromático, ilustrados sob diferentes bases de pigmentos de retina, através de vários animais. C. Neumeyer, "Das Farbensehen des Goldfishes" (Ph. D. dissertation, University of Mainz, 1986).

Anexo 6 – Diagrama genérico de um circuito combinacional



Anexo 6 – Diagrama genérico de um circuito combinacional²⁰²

²⁰² Güntzel e Nascimento, “Circuitos Combinacionais”, pt. 3. Circuitos combinacionais, 1.