



**UNIVERSIDADE ESTADUAL PAULISTA JÚLIO  
DE MESQUITA FILHO**

Faculdade de Filosofia e Ciências – Campus Marília  
Programa de Pós-Graduação em Filosofia

**NATHÁLIA CRISTINA ALVES PANTALEÃO**

**ANÁLISE DE TESES INTERNALISTAS SUBJACENTES À MODELAGEM  
COMPUTACIONAL DA MENTE EM UMA PERSPECTIVA SITUADA,  
INCOPORADA E AUTO-ORGANIZADA**

Marília  
2015

NATHÁLIA CRISTINA ALVES PANTALEÃO

**ANÁLISE DE TESES INTERNALISTAS SUBJACENTES À MODELAGEM  
COMPUTACIONAL DA MENTE EM UMA PERSPECTIVA SITUADA,  
INCOPORADA E AUTO-ORGANIZADA**

Dissertação apresentada junto ao Programa de Pós-Graduação em Filosofia da Universidade Estadual Paulista, Campus de Marília, como requisito para a obtenção do título de Mestre em Filosofia.

**Área:** Filosofia da Mente, Epistemologia e Lógica.

**Linha de pesquisa:** Ciência Cognitiva, Filosofia da Mente e Semiótica.

**Orientadora:** Dra. Mariana Claudia Broens

**Agência financiadora:** FAPESP

MARÍLIA  
2015

Pantaleão, Nathália Cristina Alves.

P197a Análise de teses internalistas subjacentes à modelagem computacional da mente em uma perspectiva situada, incorporada e auto organizada / Nathália Cristina Alves Pantaleão. – Marília, 2015.  
101 f. ; 30 cm.

Orientador: Mariana Claudia Broens.

Dissertação (Mestrado em Filosofia) – Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, 2015.

Bibliografia: f. 97-101

1. Cognição. 2. Filosofia da mente. 3. Representação mental. 4. Inteligência artificial - Filosofia. 5. Sistemas auto-organizados. I. Título.

CDD 128.2

NATHÁLIA CRISTINA ALVES PANTALEÃO

**ANÁLISE DE TESES INTERNALISTAS SUBJACENTES À MODELAGEM  
COMPUTACIONAL DA MENTE EM UMA PERSPECTIVA SITUADA,  
INCOPORADA E AUTO-ORGANIZADA**

Dissertação apresentada junto ao Programa de Pós-Graduação em Filosofia, da Faculdade de Filosofia e Ciências da Universidade Estadual Paulista Júlio de Mesquita Filho, Campus de Marília, como requisito para a obtenção do título de mestre em Filosofia, sob a orientação da Profa. Dra. Mariana Claudia Broens.

**Data de Defesa:** 08/06/2015

**BANCA EXAMINADORA:**

**Titular 1 (orientadora):** Profa. Dra. Mariana Claudia Broens (UNESP/Marília-SP).

---

**Titular 2:** Profa. Dra. Mariana Matulovic da Silva (UNESP/Marília-SP).

---

**Titular 3:** Prof. Dr. Marco Antonio Caron Ruffino (UNICAMP/Campinas-SP).

---

**Suplente interno:** Prof. Dr. Hércules de Araujo Feitosa (UNESP/Marília-SP).

---

**Suplente externo:** Prof. Dr. Kleber Bez Birolo Candiotta (PUC/Curitiba-SP).

---

*Para os meus meninos: Michael, Breno e Alexandre*

## AGRADECIMENTOS

Este trabalho é antes de tudo um pedaço significativo da minha existência. Tentei escrevê-lo com rigor conceitual, profundidade filosófica e coração nas mesmas proporções. As páginas a seguir tiveram lágrimas, sangue e sorrisos. Muitas pessoas participaram desse processo: foi um trabalho de muitas mãos.

Nesse sentido, agradeço primeiramente aos meus pais, Dalva e Zacarias, por terem me dado a oportunidade de estar nesse mundo, nesse momento e nessas condições. Agradeço aos meus irmãos Michael e Breno, por terem me escolhido. À minhas avós Anezia e Cida, mulheres fortes que me mostraram que tudo é possível com amor e paciência. À minha tia Marta por estar sempre comigo mesmo que de longe. Agradeço ao Israel por cuidar de quem eu mais amo nessa vida.

Além do meu núcleo familiar agradeço aos amigos que fiz em Marília: agradeço à Ira por ser luz e estar sempre aqui. À Lola por dividir a casa, os gatos e a companhia. À Amanda, pela resistência. Ao Tiago Brentam por compartilhar o desconforto com este mundo. Ao Alexandre por estar comigo quietinho, rindo e quando eu me senti sozinha. À Sil, pelas bolachinhas de nata da mãe dela. Ao Paulo Henrique por estar comigo até a metade do caminho. Ao João Moraes pelas leituras, por me socorrer sempre e pela alteridade. À Débora pela sinceridade, paciência e por aparecido na minha vida quando eu mais precisava. À Mari Vitti por ser princesa e muitas vezes me trazer de volta à realidade. Ao Eloi e a Leo pelos domingos. Ao Rafael Teruel pela literatura. Ao Marlon pelas conversas. À Sara por estar desde o começo. Ao João Ricardi por ser quem é. À Bruna por ter sido uma surpresa. Ao Danilo Ramos pela comida de mãe. Ao Sidney pelo yoga. À Ludmila e ao Geraldo pela paz. À Jussara pela fé. Ao Kleber pelos olhos atrás dos óculos. Agradeço ao Vini e ao Eliton por terem aparecido nos meus momentos de dificuldades (e nos de festa também). Agradeço também a todos os colegas do GAEC (Grupo Acadêmico de Estudos Cognitivos) cujas discussões contribuíram para este trabalho.

Agradeço à Edna, secretária do Departamento de Filosofia, pela calma, competência e dedicação. Aos funcionários da Seção de Pós Graduação e do Escritório de Pesquisa por todo o suporte dispensado.

Agradeço aos professores do Programa: Hércules, Luiz Henrique, Jonas, Alfredo e Marcos Alves pelas disciplinas ministradas e pela disposição de sempre tentarem sanar os meus infinitos questionamentos. Principalmente agradeço à professora Maria Eunice por me fazer (re)pensar sempre, pelas provocações e por ter participado do meu exame de qualificação. Agradeço à professora Mariana Matulovic pela leitura e sugestões sobre o meu trabalho desde a sua primeira versão.

Especialmente agradeço à professora Mariana Broens, minha orientadora que me apoiou e sempre me deu liberdade intelectual desde minha iniciação científica. Serei eternamente grata.

Por fim, agradeço ao financiamento da CAPES e da FAPESP sem o qual a realização deste trabalho não seria possível.

*Para que a gente escreve, se não para juntar nossos pedacinhos?  
(Eduardo Galeano, O livro dos abraços)*

PANTALEÃO, N. C. A. **Análise de teses internalistas subjacentes à modelagem computacional da mente em uma perspectiva situada, incorporada e auto-organizada.** Dissertação (Mestrado em Filosofia) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2015.

## RESUMO

O presente trabalho tem por objetivo analisar e problematizar teses internalistas/representacionistas presentes no projeto de modelagem mecânica da mente proposta pela Ciência Cognitiva focada em um aspecto simbólico representacionista dos processos mentais. Tais teses consideram que o comportamento inteligente tem como fundamento o conhecimento proposicional originário de representações mentais. Filósofos e cientistas cognitivos contemporâneos (Dreyfus, 1975, 1993; Searle, 1980, 1987; Brooks, 1990, 1991a, 1991b, 2002; Clark, 1997, 1999, 2003, 2008; Chemero 2007, 2009, 2012; Shapiro, 2014) têm uma visão crítica a respeito da proposta tradicional na Inteligência Artificial que segue um paradigma internalista, pois não explica satisfatoriamente muitos processos cognitivos referentes ao desenvolvimento de habilidades relacionadas a performances inteligentes, especialmente as que envolvem processos auto-organizados (Debrun, 2009; Bresciani; D'Ottaviano, 2000) em ambientes não controlados. Nesse sentido, Brooks (1990, 1991a, 1991b, 2002) argumenta à favor de um paradigma de modelagem que seja incremental, considerando a inteligência enquanto um conjunto de habilidades desenvolvidas a partir de interações agente-mundo em ambientes não controlados. Sugerimos, com Ryle (1949), que habilidades e performances inteligentes são adquiridas e desenvolvidas a partir de disposições que se efetivam/atualizam no plano da ação. Procuramos investigar quais seriam os requisitos a serem atendidos por modelos que pretendam instanciar habilidades ou performances inteligentes e analisar se os modelos da nova robótica propostos por Brooks atendem tais requisitos.

**Palavras-chave:** Internalismo *versus* externalismo. Disposições. Auto-organização. Modelos computacionais. Habilidades.

PANTALEÃO, N. C. A. **Analysis of internalist theses underlying the computational modeling of mind from a situated, embodied and self-organized perspective.** Dissertation. Faculty of Philosophy and Sciences, State University of São Paulo, Marília, 2015.

### ABSTRACT

This project aims to analyze and discuss internalist/representationalist theses present in the project of mechanical modeling of the mind proposed by Cognitive Science that focalize a symbolic and representationalist aspect of the mental processes. Such theories consider that intelligent behavior is grounded on propositional knowledge created by mental representations. Contemporary philosophers and cognitive scientists (Dreyfus, 1972, 1993, Searle, 1980, 1987, Brooks, 1990, 1991, 2002, Clark, 1997, 1999, 2003, 2008; Chemero 2007, 2009, 2012) have a critical view about what is proposed in the traditional Artificial Intelligence that follows an internalist paradigm because it doesn't explain sufficiently many cognitive processes related to the development of skills related to smart performances, especially those involving self-organized processes (Debrun, 1996; Bresciani; D'Ottaviano, 2000) in uncontrolled environments. In this sense, Brooks (1990, 1991, 2002) argues in favor of a modeling paradigm that is incremental that takes into consideration intelligence as a set of skills developed from agent-world interactions in uncontrolled environments. We suggest, with Ryle (1949), that skills and intelligent performances are acquired and developed from dispositions that become effective/updated in the plan of action. We will try to investigate the necessary requirements to be met by models that wish to instantiate skills or intelligent performances and analyze if the models proposed by the new robotic Brooks meet such requirements.

**Key-words:** Internalism *versus* externalism. Dispositions. Self-organization. Computational models. Skills.

## SUMÁRIO

<b>INTRODUÇÃO.....</b>	<b>11</b>
<b>CAPÍTULO 1 – MODELOS COMPUTACIONAIS CLÁSSICOS: HISTÓRIA, PARADIGMAS E PROBLEMAS .....</b>	<b>13</b>
<b>APRESENTAÇÃO.....</b>	<b>14</b>
1.1 AS ORIGENS DA INTELIGÊNCIA ARTIFICIAL E O PARADIGMA SIMBÓLICO .....	14
1.2 O CONCEITO DE MODELAGEM E DE INTELIGÊNCIA NA CIÊNCIA COGNITIVA .....	21
1.3 A ABORDAGEM INTERNALISTA DA MENTE: ALAN TURING E O PROJETO DE MODELAGEM MECÂNICA DE PROCESSOS MENTAIS .....	25
1.4 CRÍTICAS AO PARADIGMA SIMBÓLICO TRADICIONAL DA INTELIGÊNCIA ARTIFICIAL .....	32
<b>CAPÍTULO 2- A ABORDAGEM EXTERNALISTA DA MENTE E A MODELAGEM COMPUTACIONAL .....</b>	<b>40</b>
<b>APRESENTAÇÃO.....</b>	<b>41</b>
2.1 A TEORIA DA COGNIÇÃO INCORPORADA E SITUADA (CIS) .....	41
2.2 RYLE CRÍTICO DO REPRESENTACIONISMO .....	44
2.3 O AGENTE DISPOSICIONAL E O NEODISPOSICIONALISMO .....	48
2.4 A TEORIA DA AUTO-ORGANIZAÇÃO E AS HABILIDADES DISPOSICIONAIS.....	52
<b>CAPÍTULO 3 - A NOVA ROBÓTICA: UMA REFLEXÃO ACERCA DOS PARADIGMAS CONTEMPORÂNEOS DA IA .....</b>	<b>56</b>
<b>APRESENTAÇÃO.....</b>	<b>57</b>
3.1 O PROJETO DE BROOKS PARA UMA NOVA ROBÓTICA .....	58
3.2 A NOVA ROBÓTICA E A AUTO-ORGANIZAÇÃO: UM POSSÍVEL PARADIGMA PARA A MODELAGEM COMPUTACIONAL DA INTELIGÊNCIA? .....	69
3.3 HABILIDADES COMO UMA CAPACIDADE DE SISTEMAS COMPLEXOS: LIMITES E POSSIBILIDADES DA MODELAGEM COMPUTACIONAL DA NOVA ROBÓTICA.....	80
3.4 O PAPEL DA INFORMAÇÃO NOS PROCESSOS DE COMUNICAÇÃO E COMPLEXIFICAÇÃO .....	86
<b>CONSIDERAÇÕES FINAIS .....</b>	<b>94</b>
<b>REFERÊNCIAS.....</b>	<b>97</b>

## INTRODUÇÃO

Neste trabalho analisamos os pressupostos internalista-representacionistas da tradição filosófica em relação à explicação do comportamento inteligente e sua possível extensão na Inteligência Artificial tradicional (IA). Em seguida, apresentamos críticas dirigidas à IA tradicional (DREYFUS, 1975, 1993; SEARLE, 1980, 1987; CLARK, 1997, 1999, 2003, 2008; CHEMERO 2007, 2009, 2012) e expomos abordagens contemporâneas do conceito de *inteligência*, cujo objetivo é indicar um paradigma de modelagem computacional com uma maior plausibilidade biológica.

No que se refere aos objetivos da IA, a maioria dos pesquisadores (BROOKS, 1991; CHEMERO, 2012; TURING, 1950; MINSKY, 1991, 1995) concorda quando se afirma que seu projeto de pesquisa pretende desenvolver programas computacionais que, em princípio, possam se “comportar” ou “pensar” de maneira inteligente. Para isso, inicialmente os pesquisadores procuraram estabelecer algumas definições básicas de conceitos, como o de *programa*, *algoritmo* e o próprio conceito de *inteligência*, e técnicas de desenvolvimento de modelos computacionais. A IA busca construir modelos para simular, e talvez emular, processos inteligentes. Nesse sentido, o funcionamento do modelo constituiria um teste de uma teoria específica sobre o processo por ele simulado. Seria, justamente, a partir desse ponto que o consenso entre os pesquisadores da IA diminuiu. Isso se deve pelo fato de que algumas definições propostas por alguns pesquisadores da área enfatizam a criação de programas, outras se concentram em linguagens de programação a fim de simular os processos de pensamento, enquanto outros pesquisadores consideram relevante qualquer programa que cause comportamentos inteligentes mesmo que sejam simulações. Tendo em vista tal problemática da IA, considerando objetivo de investigar o alcance de teses internalistas e representacionistas da IA estruturamos o presente trabalho como se segue.

No *Capítulo 1* nos detivermos na clarificação de pressupostos basilares da IA, focalizando os objetivos dos primeiros pesquisadores da área e apontando as principais críticas desenvolvidas ao seu programa de pesquisa.

No *Capítulo 2*, com base nas críticas dirigidas ao projeto de modelagem da IA tradicional, analisamos uma abordagem contemporânea do conceito de *inteligência*, essencialmente pautado na Teoria da Cognição Incorporada e Situada (CLARK, 1997, 1999, 2003, 2008; HASELAGER, 2004; CHEMERO 2007, 2009, 2012; SHAPIRO, 2007, 2014) e nas teses de Gilbert Ryle (1949). Tal como aponta Shapiro (2007) a cognição é melhor analisada quando

acoplada ao plano da ação. Além do mais, tal como argumenta Chemero (2007) uma vez os agentes situados em um ambiente, o papel da representação mental no desenvolvimento de habilidades é minimizado. Apontaremos que a cognição além de poder ser abordada a partir de uma perspectiva Situada e Incorporada, envolve também processos auto-organizados. Nesse sentido, apresentaremos a Teoria da Auto-organização proposta por Debrun (2009) e como esta teoria pode contribuir para o estudo da *inteligência*.

No *Capítulo 3*, por fim, apresentamos a Nova Robótica de Rodney Brooks que propõe um projeto de modelagem computacional assentado em pressupostos da Teoria da Cognição Situada e Incorporada e que apresenta abertura para uma análise também assentada na Teoria da Auto-organização. Nesse contexto, apontaremos que a proposta de Brooks figura como uma possibilidade de outro paradigma de modelagem computacional que destoa da abordagem tradicional (tal como caracterizada no Capítulo 1). A fim de endossar a emergência de paradigmas contemporâneos de modelagem computacional apresentaremos a proposta de Beer (2003) que além de se pautar em um paradigma situado e incorporado, utiliza um esquema de redes neurais a fim de modelar comportamentos eficazes no plano da ação.

Por fim, nas *considerações finais* apontaremos para os limites e as possibilidades da proposta de Brooks aliada à Teoria da Cognição Situada e Incorporada e da Teoria da Auto-organização.

**CAPÍTULO 1 – MODELOS COMPUTACIONAIS CLÁSSICOS: HISTÓRIA,  
PARADIGMAS E PROBLEMAS**

*“In life, unlike chess, the games continues after the checkmate”  
(Asimov, 1987, p. 209).*

## APRESENTAÇÃO

Neste Capítulo apresentamos teses centrais do projeto de modelagem mecânica da mente proposto pela IA tradicional que sustentam abordagens internalistas e representacionistas de processos cognitivos. Por se tratar de um Capítulo predominantemente histórico, na **seção 1.1** realizamos uma breve exposição do surgimento da IA, com especial ênfase nos objetivos dos primeiros pesquisadores da área. Na **seção 1.2**, trouxemos à baila a discussão em torno aos conceitos de *modelagem* e de *inteligência* presentes na IA. Na **seção 1.3**, explicitamos a tentativa de Turing de modelar mecanicamente a inteligência e na **seção 1.4** apresentamos as principais críticas dirigidas ao projeto da IA tradicional.

### 1.1 AS ORIGENS DA INTELIGÊNCIA ARTIFICIAL E O PARADIGMA SIMBÓLICO

A origem da IA, enquanto área de pesquisa da Ciência Cognitiva<sup>1</sup> que visa modelar mecanicamente processos mentais, foi resultado de um empreendimento de natureza interdisciplinar. Nesse sentido, a história da IA se confunde com o caminhar de outros ramos do conhecimento, tais como a Lógica, a Matemática e a Computação. Todavia, podemos assinalar com Thagard (1998) que desde os gregos clássicos havia a aspiração de organizar logicamente os raciocínios cotidianos de modo a encontrar uma espécie de *parâmetro universal* característico dos comportamentos inteligentes.

As inquietações acerca do conceito de *inteligência* se intensificaram no decorrer do século XX, no qual as ciências da natureza passaram por revoluções que acarretaram uma mudança da nossa visão de mundo. O mesmo ocorreu com a Psicologia, que posteriormente comporia uma das áreas da Ciência Cognitiva no estudo da mente e que na década de 1930 atravessava por um período de crise: sua validade enquanto ciência era questionada e exigia-se a existência a delimitação de paradigmas, objetos de estudo e método. O impasse que permeava a Psicologia era como explicar a *psique* sem retornar ao introspeccionismo, mas também sem reduzir a mente a processos de estímulos e respostas (TEIXEIRA, 1998). Como apontado por Teixeira, tal impasse:

Tratava-se de um dilema teórico que, em última análise, colocava em cheque a possibilidade de se construir uma ciência da mente. Era preciso, de alguma forma, conciliar o reconhecimento da existência de estados internos ou

---

<sup>1</sup> Para Thagard (1998, p. 17) “A ciência cognitiva propõe que as pessoas têm *procedimentos* mentais que operam nas representações mentais para produzir pensamentos e ações.”

representações com uma proposta metodológica que afastasse a Psicologia do introspeccionismo (TEIXEIRA, 1998, p. 11).

A saída ao impasse presente na Psicologia, isto é, a aceitação de que os processos cognitivos e perceptuais resultam de representações internas, influenciou o projeto de modelagem mecânica da mente presente na, posteriormente denominada, Ciência Cognitiva. Tal projeto se inicia já em 1936 com a publicação do artigo “On Computable numbers, with an application to the Entscheidungs problem” de Alan Turing<sup>2</sup> e ganha contornos mais claros com a publicação de “Computing machinery and intelligence” em 1950. Todavia, como apontado por Teixeira (1998, p. 35) “a ideia de que a mente funciona como um computador digital e que este último poderia servir de modelo ou metáfora para conceber a mente humana iniciou-se a partir da década de 40, quando o termo ‘Inteligência Artificial’ sequer havia sido inventado”. Nesse contexto, Norbert Wiener surge como um dos pioneiros do movimento denominado *Cibernética*. Nas palavras de Wiener, o objetivo da Cibernética era responder a questão: “Como construir sistemas contendo elemento dos dois tipos, humanos e mecânicos? Creio que é somente a partir desse ponto de vista que poderemos aplicar inteligentemente as novas possibilidades de controle da cibernética” (WIENER, 1970, p. 74). Além do mais, Wiener buscava desenvolver uma teoria do controle e da comunicação, tanto em animais quanto em máquinas (WIENER, 1948/1965).

Em sentido geral, o que Wiener pretende é estabelecer meios de controle a fim de manipular as possibilidades oferecidas pela tecnologia mecânica a favor dos seres humanos. Essa proposta desencadeou uma abordagem informacional da natureza da ação, uma vez que a informação pode ser concebida como o conteúdo daquilo que um sistema pode trocar com o mundo externo para se ajustar a ele (WIENER, 1989, p. 17). Wiener considerava que existem máquinas capazes de aprender tais como, por exemplo, as máquinas de jogar damas construídas por seu grupo de pesquisa. Nas palavras de Wiener: “Há alguns anos construímos uma máquina para jogar damas e queremos construir uma outra para jogar xadrez, problema muito mais difícil porque a complexidade do xadrez é muito maior” (WIENER, 1970, p. 71). Inicialmente, as máquinas que jogam damas funcionam segundo regras dadas, mas, segundo Wiener, esse tipo de máquina parece possuir uma “personalidade” relacionada às estratégias adotadas no jogo. Em algum momento da partida essa máquina muda de personalidade e depois de algum tempo a máquina já realmente parece ser uma máquina diferente. Nesse sentido, segundo ele, a máquina “aprende” a partir de suas experiências de jogo.

---

<sup>2</sup> Exporemos de forma detalhada as teses de Turing nas seções seguintes.

A máquina, tal como caracterizada por Wiener, seria capaz de aprender quando há uma normalização da atividade (a de jogar, no caso das máquinas que jogam damas). Primeiramente, para “ensinar” a máquina a jogar damas, é necessário saber o que constitui um bom jogo de damas (suas regras, possibilidades de movimentos, o objetivo do jogo, etc). Sob esta perspectiva, podemos afirmar que há uma temporalidade envolvida nesse processo, uma vez que a máquina “desenvolve” sua “personalidade” no jogo em um tempo de experiências que ocorrem ao longo de partidas consecutivas. Como aponta Wiener: “Máquinas contribuem para a acumulação local e temporária de informação, não obstante sua organização grosseira e imperfeita se comparada com a nossa.”<sup>3</sup> (WIENER, 1989, p. 31, tradução nossa).

A possibilidade de erro também está envolvida nesse processo já que para Wiener “[...] aprender não significa acertar sempre” (WIENER, 1970, p. 72). Trata-se de trocas de informação entre a máquina e o meio. Mas como se dá essa troca? Nesse contexto, Wiener introduz o conceito de *feedback* enquanto base da Cibernética. Tal conceito pode ser entendido como a propriedade do sistema de ajustar seus comportamentos futuros a partir de informações anteriormente adquiridas (WIENER, 1989, p. 61).

Um exemplo para explicar o conceito de *feedback* é o termostato, que controla a temperatura de um ambiente a partir das informações obtidas em sua interação com tal ambiente. Uma temperatura é programada para ser considerada como parâmetro de regulação: se a temperatura do ambiente ultrapassar esse parâmetro o termostato ‘automaticamente’ regula a atividade do sistema de modo a tornar o ambiente mais fresco e, se a temperatura do ambiente for menor do que o parâmetro, o termostato desligará o sistema e assim sucessivamente.

Wiener (1989, p. 61), ao introduzir o conceito de *feedback*, rompe com a noção de *causalidade linear* (‘A’ causa ‘B’) e aponta para a ideia de *causalidade circular* (‘A’ causa ‘B’ que causa ‘A’). A causalidade circular permite o mecanismo de autoregulação, ou seja, a capacidade de ajuste a partir da troca de informações entre o sistema e o meio. O mecanismo de regulação permite, de certo modo, a autonomia de um sistema, o que inclui a possibilidade de um tipo de aprendizado já que a máquina teria a capacidade de guiar seus comportamentos futuros baseados em comportamentos passados.

Mas será que as máquinas aprendem realmente algo novo e desenvolvem uma “personalidade” como sugere Wiener? Ou esta seria apenas mais uma função dada e pré-estabelecida? Esse ‘aprendizado’ seria uma manipulação simbólica previamente programada? Essas perguntas se fundamentam no fato de que as máquinas ainda são construídas a partir de

---

<sup>3</sup>Machines also contribute to a local and temporary building up of information, notwithstanding their crude and imperfect organization compared with that of ourselves.

técnicas de engenharia e de manipulação de mecanismos artificiais. O próprio Wiener (1970, p. 73) parece perceber essa limitação ao considerar que as máquinas seriam eficazes apenas se acopladas a sistemas humanos quando observa: “[...] para que as máquinas sejam usadas de forma racional é preciso pensar em sistemas formados por elementos humanos e elementos mecânicos em trabalho conjunto e em construir os sistemas usando os dois elementos da forma que nos seja mais favorável”. Os trabalhos de Wiener influenciaram o desenvolvimento de outros projetos de pesquisa, que focalizavam majoritariamente a modelagem de sistemas inteligentes totalmente artificiais.

Posteriormente, em 1956 foi realizada em Dartmouth, nos Estados Unidos, uma conferência com duração de seis semanas que reuniu os expoentes da época em Ciência da Computação com o objetivo de, a partir da arquitetura dos computadores digitais, estabelecer os moldes para o desenvolvimento de uma ciência da mente (TEIXEIRA, 1998, p. 11). As ideias expostas nessa conferência ressoaram nas décadas seguintes: o termo **Inteligência Artificial**<sup>4</sup> disseminou-se e pesquisadores presentes na conferência começaram a associar-se em centros de pesquisas voltados ao estudo das capacidades cognitivas à luz de um modelo computacional. Este contexto impulsionou o desenvolvimento de modelos computacionais que seriam capazes de realizar satisfatoriamente uma função específica, o que implica em resolver um problema. Assim, em 1959, Newell, Shaw e Simon criam o “Solucionador Geral de Problemas (GPS)<sup>5</sup>”, um programa de computador destinado à resolução de problemas passíveis de representação simbólica, tais como resolver e provar teoremas matemáticos, jogar xadrez ou até entender o significado de um texto literário o qual, segundo os autores, não passaria do resultado de um processo de manipulação simbólica e sintática (NEWELL; SHAW; SIMON, 1963, p. 109). O objetivo consistia, então, em, através do GPS, entender o processo que efetivamente tange à resolução do problema em questão. A maneira como o estudo das resoluções de problemas acontecia era: “[...] especificar, em detalhes, programas para computadores digitais e estudar seu comportamento empiricamente, rodando os programas com variações e sob uma variedade de condições<sup>6</sup>” (NEWELL; SHAW; SIMON, 1963, p. 110, tradução nossa).

<sup>4</sup> Como apontado por Teixeira (1998, p. 12) o termo Inteligência Artificial foi cunhado por John McCarthy, um dos organizadores da conferência de 1956.

<sup>5</sup> Em 1957, Newell, Shaw e Simon já haviam desenvolvido um programa chamado *Logic Theoristic*. Como aponta Dreyfus (1975, p. 47), o Logic Theorist “utilizando uma busca heurísticamente orientada, na base do método de tentativas, provou 38 dentre 52 teoremas constantes do *Principia Mathematica* [de Alfred North Whitehead e Bertrand Arthur William Russell]. Dois anos depois, outro programa de Newell, Shaw e Simon, o Solucionador Geral de Problemas (SGP), utilizando uma análise mais sofisticada de meios e fins, resolveu o problema dos ‘canibais e os missionários’ e outros problemas de complexidade semelhante.”

<sup>6</sup>[...] specify them in detail, program them for digital computers, and study their behavior empirically by running them with a number of variations and under a variety of conditions.

Entretanto, o GPS foi construído a partir de um paradigma mecânico e procedural (tal como fez Babbage), cuja funcionalidade estava contida em instruções representadas por meio de marcas em papéis ou cartões perfurados. Tratava-se, então, de um programa puramente formal baseado em regras lógicas, mais precisamente, em cálculos sentenciais. Nas palavras dos autores:

O cálculo sentencial é um sistema matemático formalizado, consistindo de expressões construídas a partir de combinações de símbolos básicos. Cinco dessas expressões são consideradas axiomas e há regras de inferência para a geração de novos teoremas a partir dos axiomas e de outros teoremas<sup>7</sup> (NEWELL; SHAW; SIMON, 1963, p. 110, tradução nossa).<sup>8</sup>

Os processos realizados pelo GPS limitavam-se, exclusivamente, ao âmbito da matemática formal e, sendo assim, não necessitavam de algum tipo de interpretação. O comportamento do programa de Newell, Shaw e Simon era baseado nos conceitos de *problema*, *algoritmo*<sup>9</sup> e *heurística*. Para os autores, *problema* é tudo aquilo que é passível de ser resolvido: alguém teria um problema apenas se pudesse selecionar sua solução entre um conjunto de possibilidades (NEWELL; SHAW; SIMON, 1963, p. 113). Partindo do pressuposto de que para um problema existe uma solução, dado um problema específico, o GPS necessariamente deveria encontrar uma maneira de solucioná-lo<sup>10</sup>. Para Newell, Shaw e Simon a solução para problemas dados consistiria em encontrar *algoritmos* e uma *heurística* apropriados: *algoritmo* enquanto um processo delimitado e especificado a partir de regras não ambíguas e *heurística* enquanto um método simplificador que contém em si operações que resolvem o problema (NEWELL; SHAW; SIMON, 1963).

Em suma, a partir de um problema dado, o GPS testa um método (constituído por axiomas e teoremas pré-estabelecidos e armazenados em uma *lista de teoremas*) a fim de solucioná-lo, assim como exclui os métodos mal sucedidos para solucionar o problema em questão. Se o método testado falha, o problema é dividido em subproblemas, ou seja, em problemas parciais cuja probabilidade de resolução seria maior devido à amplitude de sua constituição. Esses subproblemas eram adicionados a uma *lista de subproblemas*. Se todas as

<sup>7</sup> The sentential calculus is a formalized system of mathematics, consisting of expressions built from combinations of basic symbols. Five of these expressions are taken as axioms, and there are rules of inference for generating new theorems from the axioms and from other theorems.

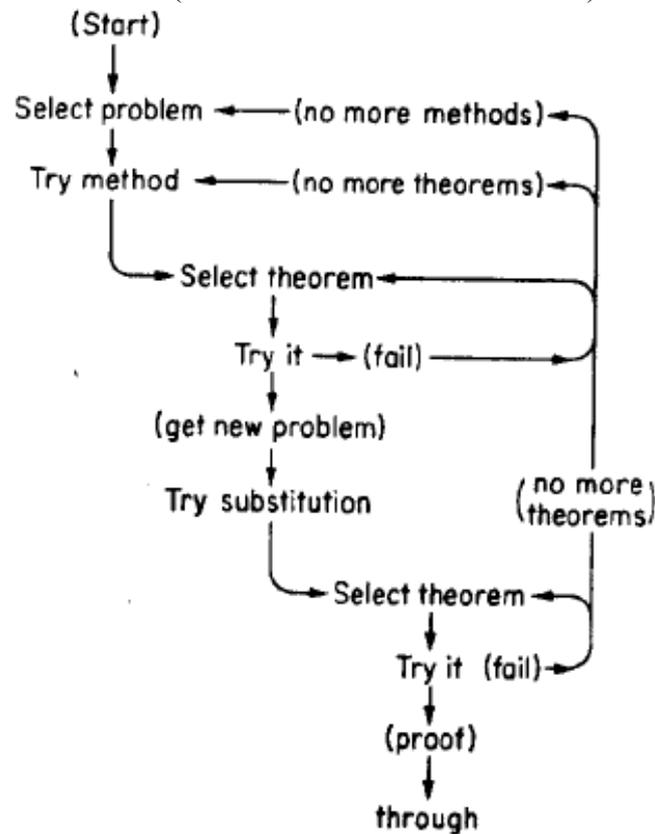
<sup>8</sup> Como apontado por Tassinari e D'Ottaviano (2012, p. 154) “Na sistematização de uma área do conhecimento, como as deduções sempre se apoiam em asserções anteriores, devemos aceitar determinadas asserções como primeiras para não cairmos em um regresso infinito; essas primeiras asserções, que aceitamos sem delas ter uma dedução, são chamadas de axiomas. A partir dos axiomas, regras de inferência estabelecem então como passar de uma asserção à outra, em deduções e demonstrações, gerando asserções chamadas de teoremas. Notemos que as regras de inferência também são argumentos válidos.”

<sup>9</sup> Tentaremos na seção 1.4, definir o conceito de *algoritmo*.

<sup>10</sup> David Hilbert (1862-1943) sentencia que em relação aos problemas matemáticos não existe o “ignorabimus”, ou seja, não existe a ignorância da resolução de um problema, pois todos estes são ou serão passíveis de resolução. Nesse sentido, Hilbert afirma: “Nós precisamos saber, e nós iremos saber”.

tentativas falham, ou seja, nenhum teorema da lista de teoremas resolve o problema ou o subproblema, a mesma rotina se repete normalmente e depois se retroalimentando. Primeiro tenta-se criar um subproblema, em seguida tenta-se resolvê-lo através de um método, se esta tentativa falha, coloca-se o subproblema na lista e como apontam Newell, Shaw e Simon (1963, p. 119, tradução nossa): “Pela natureza dos métodos, se o método de substituição é sempre bem sucedido para cada um dos subproblemas, o teorema original é provado<sup>11</sup>”. Por fim, se todos os métodos foram testados e o problema não foi solucionado, o programa seleciona o próximo subproblema e executa os mesmos passos descritos anteriormente. A figura 1 apresenta, esquematicamente, o funcionamento do GPS:

**Figura 1** - Diagrama da rotina de execução do Solucionador Geral de Problemas (*General Problem Solver – GPS*).



**Fonte:** Feigenbaum; Fheldman (1963).

O que cabe ressaltar aqui é que Newell, Shaw e Simon entendiam que o desenvolvimento do GPS representava, de fato, o pensamento humano. Para eles, o *pensar* equivaleria à manipulação de axiomas e teoremas a fim de resolver um problema artificial, no

<sup>11</sup> By the nature of the methods, if the substitution method ever succeeds a single subproblem, the original theorem is proved.

sentido de que tal problema não é legitimamente um problema do programa, mas sim um dado por fatores externos (como o programador). Além do mais, tradicionalmente, os problemas a serem resolvidos por este tipo de programa consistiam em operações matemáticas ou lógicas, embora seus proponentes afirmem que seria um programa para a solução de problemas abstratos. Essa postura revela, ainda, a presença de pressupostos internalistas, racionalistas e simbólicos desvinculados do plano da ação na medida em que os problemas solucionados por máquinas como a de Newell, Shaw e Simon são limitados à um contexto específico no qual a manipulação simbólica regida por regras específicas é suficiente.

O que podemos ressaltar aqui, assim como faz Teixeira (1998, p.12) é que a maioria dos pesquisadores da IA, desde o seu surgimento, planejavam:

[...] *simular* efetivamente processos mentais humanos e usar o computador para fundar uma ciência da mente. Para tanto, era preciso expandir a aplicação do modelo computacional para outros domínios além da simulação do raciocínio como acontecia com as máquinas de jogar xadrez ou demonstrar teoremas da lógica.

Contemporaneamente, pesquisadores como Dennett (1995) apontam que mais do que simular processos mentais humanos, os modelos computacionais da IA são experiências de pensamento reais que exemplificam teorias explicativas de como a mente funciona. Segundo Dennett, os modelos construídos segundo o paradigma da IA, carregam em si a demonstração efetiva de como determinado processo mental ocorre. Por exemplo, um robô que seja capaz de desviar de obstáculos que lhe são impostos, carrega em si um arcabouço de pressupostos de como o processo de *reconhecer* obstáculos e *desviar* deles ocorre no plano humano (que é considerado o natural). No contexto da modelagem computacional as hipóteses poderiam ser refutadas concretamente. Poderíamos *testar*, *comprovar* ou *refutar* as hipóteses acerca do conceito de inteligência por meio do funcionamento e do desempenho dos modelos computacionais nos quais são instanciados os pressupostos que dizem respeito ao conceito. Por exemplo, se um pesquisador defende que calcular é inteligência e constrói um modelo que calcula satisfatoriamente, então, sua hipótese foi corroborada por meio do modelo computacional. Podemos dizer que, segundo Dennett, um modelo apenas se diferenciaria do orgânico devido à sua constituição física e sendo feito de outra matéria e tendo as mesmas capacidades humanas, constituiria uma ferramenta explicativa.

A IA trouxe a tona a problematização sobre a natureza da mente e sobre o próprio conceito de inteligência. Se a IA não conseguiu, satisfatoriamente, construir máquinas

inteligentes, ela trouxe à baila a reflexão acerca do significado do conceito de inteligência e de outros conceitos amplamente empregados na Filosofia da Mente e na Ciência Cognitiva. O conceito de *inteligência* é basilar na Ciência Cognitiva e o modelo, nesse contexto, constitui uma ferramenta explicativa. Tendo em vista tal problemática trazida pela IA, abordaremos na seção seguinte os conceitos de *inteligência* e de *modelagem*.

## 1.2 O CONCEITO DE MODELAGEM E DE INTELIGÊNCIA NA CIÊNCIA COGNITIVA

O ramo da Ciência Cognitiva denominado Inteligência Artificial surgiu do projeto de pesquisadores de diversas áreas que se propunham a entender e reproduzir comportamentos humanos considerados inteligentes. Mas em que consiste a *inteligência*? O que se quer dizer quando se diz que alguém (ou um sistema) *é inteligente*? Como apontamos na seção anterior, para os primeiros pesquisadores da IA, a inteligência consistia na atividade de resolver satisfatoriamente problemas. Nesse sentido, é inteligente o ser capaz de manipular e empregar certos conhecimentos (regras, teoremas, etc.) de modo a alcançar uma solução satisfatória de um problema a partir de um conjunto de soluções possíveis. Todavia, essa é apenas uma parcela da amplitude do conceito de *inteligência* que ainda não é bem definido. Como aponta Murphy (2000, p. 15, tradução nossa):

[A] Inteligência Artificial em geral não tem definições comumente aceitas. Um dos primeiros textos da IA defini-a como ‘o estudo de ideias que possibilitam computadores inteligentes’, o que parecia evitar a questão [o que é ser inteligente?]. Um texto recente é mais específico, ‘IA é a tentativa de que o computador faça coisas que, no momento, pessoas fazem melhor’. Esta definição é interessante porque implica que, sendo a tarefa realizada com sucesso por um computador, então a técnica que tornou isso possível já não é mais IA, mas algo mundano [do mundo]<sup>12</sup>.

Podemos apontar que o conceito de *inteligência* também diz respeito ao plano da ação e não só ao da manipulação simbólica, cuja relevância foi maximizada pelos primeiros cientistas cognitivos. A satisfatoriedade com a qual os problemas são resolvidos diz respeito à resolução natural e humana que serve de modelo para a representação de tais problemas. Nesse sentido, o paradigma da inteligência adotado pela IA focaliza e privilegia certas

---

<sup>12</sup> Artificial Intelligence has no commonly accepted definitions. One the first textbooks on AI defined it as ‘the study of ideas that enable computers to be intelligent,’ which seemed to beg the question. A later textbook was more specific, ‘AI is the attempt to get the computer to do things that, for the moment, people are better at’. This definition is interesting because it implies that once a task is performed successfully by computer, then the technique that made it possible is no longer AI, but mundane.

capacidades de seres humanos e tais capacidades servem de modelo para a construção de sistemas artificiais. O que vale ser ressaltado aqui é que os pesquisadores da IA mantiveram (e mantêm) a característica de *inteligência* no processo de abstração para a construção de modelos. A modelagem consiste justamente no desenvolvimento de modelos, ou seja, de uma forma que representa um dado fenômeno.

Assim, podemos dizer que o método de modelagem adotado pelos pesquisadores da IA revela suas concepções pessoais porque, dentre tantas características humanas, privilegiaram a *inteligência* no processo de modelagem computacional? Uma resposta a esta questão diz respeito à hipótese de que a inteligência, ao ser considerada um processo de manipulação simbólica a fim de solucionar um problema, seria a capacidade humana mais relevante e com maior plausibilidade de modelagem. Podemos dizer que tal hipótese se fundamenta em técnicas de programação lógica e mecanização física vigente, pois estas impossibilitam, em certo sentido, a modelagem de outros tipos de características humanas, como as relacionadas a emoções, por exemplo, e porque os cientistas cognitivos consideram, efetivamente, a inteligência simbólica a expressão específica da condição humana.

A postura de que a inteligência diz respeito à manipulação simbólica permeia não apenas o ramo da Ciência Cognitiva, mas também outros ramos das Ciências Humanas, como a Psicologia e a Filosofia da Linguagem. Vemos que tal postura parece perpetuar e influenciar a maioria dos sistemas explicativos em relação ao conceito de *inteligência*. Nesse sentido, o psicólogo Robert Sterbenberg (1986, p. 51) afirma que: “[...] a inteligência envolve a capacidade de solucionar problemas abstratos”. Esta caracterização da inteligência se estende ao projeto de modelagem computacional da Ciência Cognitiva, mas é curioso notar que esse tipo de caracterização já estava presente em pensadores racionalistas como René Descartes. Segundo Descartes, a inteligência é uma característica que diz respeito à mente enquanto substância pensante (*res cogitans*). Nesse sentido, a substância corpórea (*res extensa*) está submetida a leis físicas, mas estaria, segundo Descartes, causalmente relacionada à mente. Nas palavras do filósofo: “Considerava-me, inicialmente, como provido de rosto, mãos, braços e toda essa máquina composta de ossos e carnes, tal como um cadáver, a qual eu designava corpo” (DESCARTES, 1988, p. 25).

Por outro lado, a mente não se sujeitaria, para Descartes, às restrições da corporeidade, pois não se encontra no espaço, suas funções não estão sujeitas a leis mecânicas e não são observáveis: seu desenvolvimento é privado e se caracteriza por uma experiência individual e introspectiva. Em relação à mente ou substância pensante, Descartes (1988, p. 25-26) afirma, utilizando a primeira pessoa, que:

[...] é um atributo que me pertence; só ele não pode ser separado de mim. *Eu sou, eu existo*: isto é certo; mas por quanto tempo? A saber, por todo o tempo em que penso; pois poderia, talvez, ocorrer que, se eu deixasse de pensar, deixaria ao mesmo tempo de ser ou de existir. Nada admito agora que não seja necessariamente verdadeiro: nada sou, pois, falando precisamente, senão uma coisa que pensa, isto é, um espírito, um entendimento ou uma razão, que são termos cuja significação me era anteriormente desconhecida.

A essência do ser humano é assentada na faculdade de pensar e esta, por sua vez, é fundamentada metafisicamente por vias racionais. Em outra passagem Descartes deixa claro a natureza do ser: “compreendi por aí que era uma substância cuja essência ou natureza consiste apenas no pensar e que não necessita de nenhum lugar nem depende de qualquer coisa material” (DESCARTES, 1987, p. 47). Assim sendo, o pensamento constitui a essência do ser na medida em que o torna capaz de atribuir sentido às coisas do mundo através da reflexão interna. Tal como apontado por Khalifa (1996, p. 7-8), “Os ‘dualistas’ postulavam que a inteligência era uma faculdade exclusiva dos seres regrados por uma substância imaterial, um espírito, do qual uma das mais claras manifestações no mundo era o comando da fala [...]”. Pelo viés dualista substancial, a inteligência não diria respeito a um substrato físico, mas imaterial ou mental relacionado à capacidade de representar e de manipular regras lógicas, tese com a qual concordam, *grosso modo*, os cientistas cognitivos. Entretanto, os cientistas cognitivos contemporâneos defendem a tese de que comportamentos inteligentes podem ser simulados por sistemas artificiais, como fazem, por exemplo, Newell; Shaw; Simon (1963), Putnam (1967/1975) e Minsky (1991) e, assim sendo, a inteligência não mais seria uma propriedade exclusivamente humana relativa a uma entidade metafísica.

Como já observamos, embora o conceito de inteligência seja de grande importância na Ciência Cognitiva, não se trata de um conceito bem definido. Nesse sentido, a construção de modelos se apresenta também como método e ferramenta explicativa de conceitos nebulosos como o de *inteligência*. Trata-se da tentativa de explicar processos complexos através de mecanismos fundamentados na demonstração e na dedução. Desse modo, por meio do funcionamento do modelo, o próprio conceito em questão se tornaria, em certo sentido, auto-evidente. Como apontado por Dupuy (1996, p. 21): “O que o homem faz, ele pode conhecê-lo racionalmente, de maneira demonstrativa e dedutiva, apesar da finitude de seu entendimento”. Ou seja, a partir da construção de modelos, o ser humano conhece e se reconhece. Mas o que é um **modelo**? Ainda segundo Dupuy:

O modelo é como uma forma abstrata que vem encarnar-se ou realizar-se nos fenômenos. [...] Campos muito diferentes da realidade fenomenal, como

a hidrodinâmica e a eletricidade, a luz e as vibrações sonoras podem ser representados por meio de modelos idênticos, o que estabelece entre eles uma relação de equivalência. O modelo é a classe de equivalência correspondente. Isso lhe confere uma posição de sobrepujança, como uma Ideia platônica cujo real não é mais que a pálida cópia. Mas o modelo, na ciência, é o que o homem faz. Ai está como se dá a sobreposição do imitante e do imitado. O modelo científico é uma imitação humana da natureza que o cientista logo toma como “modelo” – no sentido comum – desta (DUPUY, 1996, p. 23).

Inicialmente o modelo diz respeito à construção de mecanismos artificiais de objetos e características que, dada sua importância, merecem ser imitadas (ou, no caso, modeladas). O modelo, em certo sentido, possibilita que um recorte da realidade, de acordo com um critério de relevância, seja controlado e suas variáveis minimamente previstas mesmo que seu comportamento não esteja contido inicialmente no modelo. Nas palavras de Dupuy (1996, p. 24), “O modelo abstrai da realidade fenomenal o sistema das relações funcionais consideradas por ele as únicas pertinentes, pondo, por assim dizer, entre parênteses tudo o que não depende desse sistema [...]”. Desse modo, por ser a representação de uma parte específica e delimitada de fenômenos, muitas vezes o modelo acaba se distanciando da realidade. Em outras palavras, por não conter a representação da totalidade dos fenômenos da realidade, o modelo se emancipa, em um certo sentido, dessa realidade, embora ainda contenha em si as representações básicas que o tornam um *modelo*. Como aponta ainda Dupuy:

Os modelos têm, pois, uma vida própria, uma dinâmica autônoma, desligada da realidade fenomenal. Em sua atividade modelizadora, o cientista projeta sua mente no mundo das coisas. Os enormes sucessos da modelização científica são como um testemunho de que a mente é ao mesmo tempo distinta da matéria e adequada a ela. Isso não deixa de ter seus perigos. O modelo é tão mais puro, tão melhor controlável do que o mundo dos fenômenos que existe o risco de que ele se torne objeto exclusivo da atenção do cientista (DUPUY, 1996, p. 25)

O modelo apresenta-se como uma ferramenta de controle explicativo e preditivo de maior eficiência, rapidez e elegância que o mundo fenomenal não possui. Em um certo sentido, o ser humano conhece quando consegue organizar o mundo fenomenal. Como afirma Dupuy (1996, p. 27): “Conhecer é produzir um modelo do fenômeno e efetuar sobre ele manipulações ordenadas. Todo conhecimento é **reprodução, representação, repetição, simulação**” (destaque nosso). Nesse sentido, o modelo é uma estrutura de representação de um objeto ou, mais especificamente, de um recorte do mundo fenomenal que serve de explicação para um fenômeno específico. Como aponta Minsky (1995, p. 1) “Nós usamos o termo *modelo* nesse sentido: *Para um observador B, um objeto A\* é um modelo de um objeto*

*A na medida em que B pode utilizar A\* para responder questões que lhe interessam sobre A.*<sup>13</sup>”.

Quando se trata de um modelo computacional, consideramos que esse tipo de modelo é uma estrutura lógica de representação de um processo ou capacidade cognitiva de um sistema, cuja eficácia se revela através do desempenho de uma função computável. Uma função é computável quando existem regras definidas e explícitas que, ao serem seguidas, são capazes de calcular o valor para quaisquer argumentos dados (BOLOS *et al.*, 2012). Alan Turing foi o responsável por desenvolver a máquina teórica que acabou por formalizar a noção de computação e, conseqüentemente, a de algoritmo, conceitos chave para o desenvolvimento dos modelos da IA. Abordaremos as teses de Turing na seção seguinte.

### 1.3 A ABORDAGEM INTERNALISTA DA MENTE: ALAN TURING E O PROJETO DE MODELAGEM MECÂNICA DE PROCESSOS MENTAIS

Desde a modernidade clássica, as tentativas da tradição internalista/intelectualista de oferecer respostas para interrogações sobre a natureza da inteligência pareceram satisfatórias por apresentarem uma justificação aparentemente sólida (e supostamente evidente) a saber, o próprio ser humano e suas estruturas racionais internas com uma ontologia própria. Leclerc (2003) considera como exemplo claro do internalismo a tradição cartesiana e a caracteriza como uma perspectiva segundo a qual:

Nenhuma característica mental (estado, evento, ato) depende da existência de qualquer outra coisa fora do sujeito. [...] Nesta perspectiva, a posse, por um sujeito, ou a atribuição a um sujeito, de uma propriedade psicológica não depende em nada de existência de qualquer coisa “fora da mente” presente no ambiente natural e social (2003, p. 141).

Leclerc (2003) aponta ainda que crenças, intenções, desejos, imaginações, dentre outros, são exemplos típicos de propriedades psicológicas. Assim, de acordo com o internalismo, um indivíduo possuiria uma propriedade psicológica se, e somente se, tal propriedade for intrínseca a ele (LAU; DEUTSC, 2010). Como endossado por Bizarro (1999, p. 2): “Uma posição internalista acerca dos conteúdos mentais defende que estes devem ser

---

<sup>13</sup> We use the term *model* in this sense: *To an observer B, an object A\* is a model of an object A to the extent that B can use A\* to answer questions that interest him about A*

individualizados recorrendo principalmente à informação acerca do indivíduo que os protagoniza, informações acerca das características ‘internas’ desse indivíduo”.

Na perspectiva internalista, os sentidos, ao apreenderem o meio externo, são passíveis de erros e ilusões (tais como miragens e ilusões de ótica) e por isso alguma entidade mental interna deveria agir como ferramenta epistêmica para corrigir tais erros ou ruídos, criando representações adequadas dos objetos da experiência fenomênica. Assim procedendo, o indivíduo cognoscente, em tese, poderia conhecer os objetos através de ferramentas intelectuais sem a presença de erros, ilusões ou ruídos de modo geral. Como apontado por Haselager (2004, p. 106): “As representações são postuladas para atuar como intermediárias entre a percepção e a ação, especificando informações sobre o mundo, frequentemente de forma independente do contexto – e da ação”. Em outras palavras, a representação mental tomaria o lugar de elementos externos (carregando seus conteúdos), faria a intermediação entre estes e o indivíduo cognoscente e, por fim, desencadearia um dado comportamento (HASELAGER, 2004). Desse modo, a representação mental se configuraria como uma ferramenta explicativa e satisfatória para a ação que passa a ser considerada enquanto um processo interno do indivíduo cognoscente. A representação mental ocupa um lugar central na Ciência Cognitiva tradicional e tal papel é fortalecido pelo desenvolvimento de modelos computacionais (baseados em representações) que apresentam comportamentos de seres humanos considerados inteligentes (HASELAGER, 2004).

Nesse viés, os pressupostos do internalismo parecem perdurar nos modelos computacionais propostos pela IA, inicialmente tendo como expoente Alan Turing. Turing formula uma máquina universal e abstrata na tentativa de resolver o problema lógico da *decisão*, postulado por Hilbert. O problema da decisão consiste em provar a existência de um procedimento geral e efetivo que, dada uma fórmula qualquer do cálculo de predicados, permita determinar se essa fórmula é logicamente demonstrável ou não (DUPUY, 1994, p. 28). O que está implicado no problema de Hilbert é o conceito de *algoritmo*. Ou seja, a resolução do problema da decisão implica a existência de um procedimento sistemático e mecânico que, a partir de axiomas e regras de inferência, determine se um enunciado dado é logicamente demonstrável.

Nesse contexto, Turing propõe uma máquina abstrata, cujos princípios são apresentados no artigo *On Computable Numbers, With an Application to the Entscheidungsproblem* de 1936. Segundo Turing, a máquina seria composta por uma fita dividida em células; a fita é considerada como finita, mas quadrados brancos podem ser adicionados às extremidades (ou seja, a fita é *potencialmente infinita*). Além do mais, a máquina seria

composta por um dispositivo de leitura e uma tabela de instruções. O dispositivo de leitura é capaz de ler, escrever, apagar e movimentar-se por meio dos quadrados da fita. As ações desse dispositivo são determinadas pela tabela de instruções (o programa) da máquina em relação ao seu estado atual. Assim, a máquina sempre está em algum estado entre infinitos estados, mas ela trabalha com um aparato finito de instruções. A observação do estado corrente da máquina gera uma operação e um novo ou o mesmo estado de acordo com as regras contidas na tabela de instruções. Como apontam Carnielli e Epstein, as operações de uma máquina de Turing consistem nas seguintes instruções:

- a. Escreva o símbolo 1 no quadrado observado: 1
- b. Apague qualquer símbolo que apareça no quadrado observado: 0
- c. Mova um quadrado para a direita do quadrado observado:  $D$
- d. Mova um quadrado para a esquerda do quadrado observado:  $E$

Assim, uma instrução completa para a máquina consiste de uma *quádrupla*  
 $(q_i, S, Op, q_j)$

Onde  $q_i$  é o estado corrente,  $S \in \{0, 1\}$  é o símbolo corrente (lembre-se que '0' significa simplesmente que a fita está em branco),  $Op \in \{1, 0, D, E\}$  é uma das operações acima, e  $q_j$  é o novo estado. Note que permitimos ambos  $(q_i, 0, 0, q_j)$  e  $(q_i, 1, 1, q_j)$ , que querem dizer que a máquina apenas muda de estado" (CARNIELLI; EPSTEIN, 2005, p. 109).

A máquina idealizada por Turing não faria nada além do que sua tabela de instruções pré-estabelecida determinasse. Cada quadrado da máquina poderia ter dois estados ( $q_i$ ): conter uma marca ou estar em branco. Para a manipulação de cada quadrado existe um conjunto de operações pré-estabelecidos ( $S$ ). Tal conjunto contém as possibilidades de ação da máquina: escrever, apagar, mover para a direita e mover para a esquerda. A partir desse conjunto a máquina teria outro estado ( $q_j$ ). Para tanto, a tabela de instruções da máquina nunca apresenta conflitos lógicos e quando não há uma instrução possível de ser seguida, a máquina para. Em outras palavras, por meio de instruções codificadas (em uma linguagem mecânica e lógica), uma dada máquina de Turing realiza uma função particular cuja representação se encontra nas marcas nos quadrados da fita. Trata-se, portanto, de um conjunto de manipulações simbólicas.

Dessa maneira, a máquina de Turing permite determinar o que é *computável*, ou seja, permite determinar se para  $x$  existe um número finito de passos que o representam. O problema decorrente da formalização de Turing consiste justamente em determinar se existe um procedimento mecânico que faça a máquina parar após um número finito de passos e assim realizar uma computação. Como apontado por Dupuy (1995, p. 34), Turing mostra que o problema da parada é equivalente ao problema da decisão de Hilbert e demonstra logicamente que não existe uma máquina que, munida de um procedimento mecânico seja

capaz de determinar a parada. Assim, Turing responde negativamente à questão levantada por Hilbert.

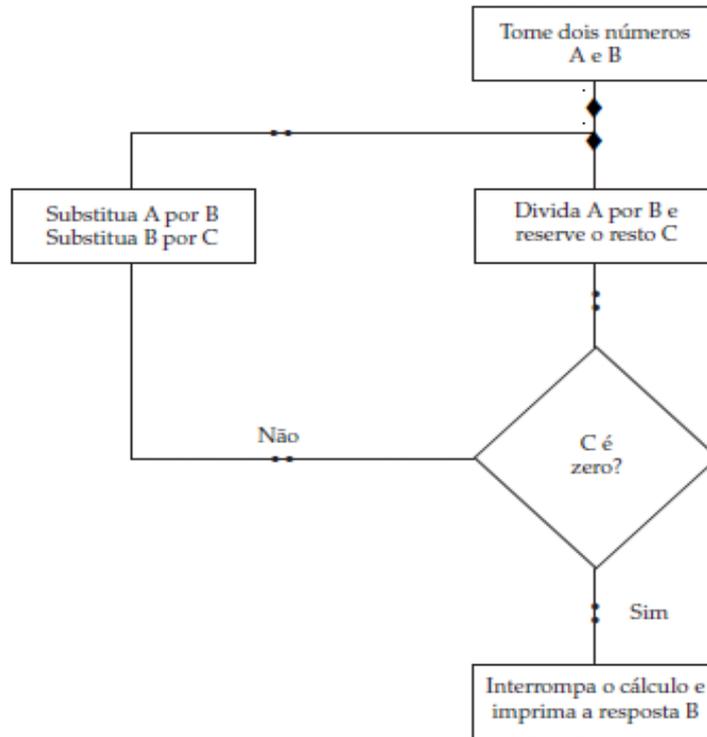
Todavia, a proposta decorrente do projeto de Turing é que a mente humana pode ser concebida como uma máquina de Turing, ou seja, a máquina pode ser considerada enquanto um modelo mecânico de mente. Como apontado por Dupuy (1995, p. 35), segundo Turing, a mente pode ser considerada uma máquina de Turing que opera a partir de enunciados internos análogos a enunciados lógicos-matemáticos. Os símbolos que são escritos na fita da máquina possuem três perspectivas ontológicas: são materiais e sujeitos às leis da física, possuem uma forma e são regidos por regras sintáticas e são dotados de um sentido e, portanto, têm um valor semântico. Segundo Dupuy (1995, p. 36), “O paralelismo entre sintaxe e semântica é, por seu lado, garantido pelos teoremas lógicos de consistência e de completude: os processos mecânicos que realizam as regras sintáticas mantem a coerência interna das representações simbólicas, bem como sua adequação ao que elas representam”. Assim, o modelo opera com dois níveis de análise: um sintático e um semântico, sendo que o nível sintático opera apenas com regras lógicas e o nível semântico opera com “significados”. O significado decorre da interpretação das fórmulas, regras e outros componentes sintáticos.

Além do mais, como apontado por Teixeira (1998), Turing formalizou a noção de *algoritmo* através da noção de máquina, na medida em que o mecanismo realiza os processos tais como especificados pelas regras lógicas e pela formalização da execução de funções procedurais (como já havia sido proposto por Babbage em 1888). O conceito de algoritmo não é facilmente definível, mas uma definição operacional consiste em considerá-lo como: “[...] um procedimento *mecânico* para a solução de um problema matemático, um procedimento que envolve uma representação bem definida do problema em questão e um conjunto de regras que levem a sua solução” (TEIXEIRA, 2004, p. 85). Um algoritmo é um procedimento geral, entendido como um processo cuja execução é totalmente especificada. Como apontam Carnielli e Epstein (2005, p. 98):

Isso significa, entre outras coisas, que devemos saber expressar as instruções para a execução do processo num texto *finitamente longo*. Não há espaço para a prática da imaginação criativa do executor. Ele deve trabalhar servilmente, de acordo com instruções que lhe são dadas e que determinam tudo nos menores detalhes.

Assim, um algoritmo é caracterizado por ser geral, exato, elementar e direcionado (para a resolução de um problema específico). O fluxograma a seguir apresenta a estrutura de um algoritmo.

**Figura 2** - Fluxograma do algoritmo de Euclides.



Fonte: TEIXEIRA (2004, p. 22), adaptado de Penrose (1989, p. 32).

Em seguida temos o mesmo algoritmo escrito em pseudocódigo que não diz respeito a nenhuma linguagem de programação, mas mostra a estrutura geral do algoritmo:

**Figura 3** - Construção em pseudocódigo do Teorema de Euclides.

```

algoritmo "mdc"
var
contador: inteiro
n1,n2: inteiro
maior, menor,mdc:inteiro
inicio
Escreva ("Informe um numero inteiro: ")
Leia (n1)
Escreva ("Informe um numero inteiro: ")
leia (n2)
Se n1>n2 entao
maior <- n1
Senao
maior <- n2
FimSe
Para contador <- 1 ate maior FACA
Se ((n1 mod contador = 0) e (n2 mod contador = 0)) Entao
mdc <- contador
FimSe
FimPara
Escreval ("O máximo divisor comum e: ",mdc)
finalgoritmo
  
```

Fonte: Elaborado pela autora

O algoritmo apresentado na figura 2 é conhecido como *algoritmo de Euclides* e trata-se de um procedimento para encontrar o máximo divisor comum entre dois números inteiros diferentes de zero. A figura 3 apresenta o mesmo algoritmo no qual podemos notar que as regras para o procedimento são previamente estabelecidas e apenas uma regra específica é executada em cada passo. Apenas quando a condição é satisfeita (**Se...então**<sup>14</sup>), ou seja, a entrada faz com que tal condição seja verdadeira, o algoritmo continua seu processo. A figura 3 mostra a característica procedural deste tipo de algoritmo. Trata-se de testar, linha por linha, os comandos do algoritmo até que todas as condições lógicas, tal como estabelecidas, sejam satisfeitas.

O principal objetivo da IA tradicional seria, assim, investigar a hipótese levantada por Turing (1950) segundo a qual a mente, enquanto instanciadora de pensamentos, poderia funcionar tal como um algoritmo que dita as regras para a execução de uma função (no caso a função de pensar). O algoritmo, enquanto determinante de uma função operacional da máquina, seria a representação interna, mecânica e artificial de uma *capacidade* real dos seres vivos. Assim, a IA tradicional está voltada para a construção de modelos pautados em sistemas representacionais eficientes. A contribuição de Turing diz respeito à hipótese de que funções desempenhadas pela mente (tal como calcular) podem ser formalizadas mecanicamente de modo que máquinas artificiais sejam capazes de realizá-las.

Turing (1950) assinala, ainda, a grande dificuldade de responder a questão “Podem as máquinas pensar?” decorrente do problema de definir os próprios conceitos de “máquina” e “pensamento”. Para tanto, o autor propõe, então, uma nova formulação da questão em termos de um jogo chamado “jogo da imitação” (TURING, 1950, p. 433). O jogo da imitação consiste de três participantes, um homem (A), uma mulher (B) e um interrogador (C). O objetivo do jogo é que o interrogador, separado dos outros participantes, descubra quem é o homem e quem é a mulher baseado apenas nas respostas dadas às questões. Mas, como Turing aponta (1950, p. 434):

<sup>14</sup> Algoritmos do tipo condicional (se...então...) obedece a seguinte tabela, tomando A como um comando-antecedente e B como uma execução-consequente:

Se A for falso	E B for falso	Então A->B é verdadeiro
Se A for falso	E B for verdadeiro	Então A->B é verdadeiro
<b>Se A for verdadeiro</b>	<b>E B for falso</b>	<b>Então A-&gt;B é falso</b>
Se A for verdadeiro	E B for verdadeiro	Então A->B é verdadeiro

Nota-se que a operação será apenas falsa (não acontecerá) quando o primeiro antecedente for verdadeiro e o segundo falso.

O que acontecerá quando uma máquina ocupar o lugar de A [o homem] nesse jogo? Será que o interrogador decidirá erroneamente com a mesma frequência, quando o jogo é jogado dessa forma, do que quando o fazia ao tempo em que o jogo era jogado entre um homem e uma mulher? Estas questões substituem a pergunta original “Podem as máquinas pensar?”.

O “jogo da imitação<sup>15</sup>” serviria como um teste no qual, se um computador puder se fazer passar por um ser humano, imitando-o satisfatoriamente de modo a enganar o interrogador, ele teria que ser reconhecido como inteligente. Nesse caso, a resposta à pergunta inicial sobre a possibilidade de máquinas pensarem seria afirmativa. Mas a Máquina de Turing constitui um bom modelo para criar máquinas efetivamente pensantes? Se equiparmos, tal como faz a IA tradicional, pensamento à execução de funções, a resposta a essa pergunta também é afirmativa. Para Turing (1950), a teoria da computabilidade forneceria bases para o entendimento do pensamento humano, já que para ele “pensar é computar”. Mas será esta máxima correta? Pensar consiste, então, em somente computar? Ou como aponta Fetzer (2000, p. 53): “Como podemos saber se os seres humanos são sistemas formais automatizados?”.

Endossando a postura da IA tradicional, o Funcionalismo sustenta que estados mentais são definidos e caracterizados pelo *papel funcional* que desempenham em um sistema. Dentre os funcionalistas, Putnam (1967/1975) adota<sup>16</sup> a perspectiva do argumento da múltipla realizabilidade, segundo o qual as mesmas funções podem ser instanciadas em diferentes estruturas físicas. Assim, quando consideramos as capacidades mentais dos organismos, sistemas artificiais não orgânicos também poderiam, em tese, instanciar capacidades semelhantes. Nas palavras de Putnam (1967/1975, p. 437):

[...] se o projeto de encontrar leis psicológicas que não sejam específicas de uma espécie — isto é, de encontrar uma teoria psicológica comum a diferentes espécies — alguma vez for bem-sucedido, então trará na sua esteira um delineamento do tipo de organização funcional que é necessária e suficiente para um dado estado psicológico, como também uma definição precisa da noção de “estado psicológico”.<sup>17</sup>

<sup>15</sup> Atualmente, o prêmio Loebner premia a máquina que melhor imita um ser humano, ou seja, uma máquina que passaria no teste de Turing. Nenhum projeto de máquina conseguiu atingir o primeiro lugar até hoje.

<sup>16</sup> Atualmente Putnam não sustenta mais uma posição funcionalista, no sentido de que computadores digitais são os melhores modelos para o funcionamento da mente.

<sup>17</sup>[...] if the program of finding psychological laws that are not species-specific –i.e., of finding a normal form for psychological theories of different species – ever succeeds, then it bring in its wake a delineation of the kind of functional organization that is necessary and sufficient for a given psychological state, as well as a precise definition of the notion ‘psychological state’.

Na concepção tradicional da IA, os mecanismos funcionais carregariam o cerne do comportamento inteligente enquanto execuções de funções. O substrato físico apenas desempenharia o papel secundário de propiciar condições adequadas para que as funções possam ser implementadas. Por exemplo, a função do *hardware* de um computador seria servir de base para a implementação de um *software*. Nesse sentido, o funcionalismo pode ser descrito através de uma máquina de Turing (potencialmente um computador digital). Como apontado por Maloney (1999, p.333): “Para o funcionalismo a mente, como um computador, pode processar informações e comportamentos simplesmente através da implementação de uma máquina de Turing”. Alguns filósofos e cientistas cognitivos criticaram os pressupostos da modelagem computacional que se pauta em um paradigma internalista e posteriormente funcionalista. Abordaremos tais críticas na seção seguinte.

#### 1.4 CRÍTICAS AO PARADIGMA SIMBÓLICO TRADICIONAL DA INTELIGÊNCIA ARTIFICIAL

Marvin Minsky teceu algumas críticas ao paradigma internalista e simbólico presente na modelagem computacional. Minsky (1991, p. 35-36, tradução nossa) afirma que “[este paradigma simbólico] não é o melhor caminho para representar o conhecimento ou para resolver problemas [...]”<sup>18</sup>. E continua afirmando que:

[...] sistemas lógicos puramente simbólicos são deficientes para representar as conexões heurísticas mais relevantes entre as coisas [...]. A versatilidade de que precisamos só pode ser encontrada em arquiteturas de grande escala que possam explorar e gerenciar as vantagens de vários tipos de representações ao mesmo tempo. Cada um pode ser utilizado para superar as deficiências dos outros. Para cumprir tal tarefa, cada tipo de representação do conhecimento ou inferência formalmente claro (*neat*) deve ser complementado com algum tipo de mecanismo bagunçado (*scruffy*) que pode incorporar as conexões heurísticas do conhecimento propriamente dito e o que nós esperamos fazer com ele (MINSKY, 1991, p. 36, tradução nossa).<sup>19</sup>

Minsky aponta para a necessidade de outro tipo de paradigma na modelagem computacional de capacidades cognitivas consideradas inteligentes. Tradicionalmente, os

<sup>18</sup>[...] no one best way to represent knowledge or to solve problems [...].

<sup>19</sup>[...] purely symbolic logical systems are inherently deficient in abilities to represent the all-important *heuristic connections* between things [...]. The versatility that we need can be found only in larger-scale architectures that can exploit and manage the advantages of several types of representations at the same time. Then, each can be used to overcome the deficiencies of the others. To accomplish this task, each formally neat type of knowledge representation or inference must be complemented with some scruffier kind of machinery that can embody the heuristic connections between the knowledge itself and what we hope to do with it.

pesquisadores da IA procuravam construir máquinas para resolver problemas a partir de uma estratégia *top-down*. Tal estratégia consiste em extrair a partir do nível do senso comum processos que poderiam desencadear em ações inteligentes (MINSKY, 1991, p. 36). Em outras palavras, a estratégia *top-down*, a visão geral do sistema é estabelecida tendo como ponto de partida uma instância final que desencadeia em instancias iniciais, como uma engenharia reversa. Ainda segundo Minsky (1991), os programas com base na análise *top-down* têm tido êxito em áreas técnicas e especializadas, devido ao amadurecimento das técnicas de representação do conhecimento. Mas quando aplicadas a solução do problema do senso comum, as técnicas *top-down* não apresentam muito sucesso. Como apontado por Minsky (1991, p. 36, tradução nossa):

[...] os objetos e atividades da vida cotidiana são demasiadamente variados para serem precisamente descritos por definições e deduções lógicas. A realidade do senso comum é muito desordenada para ser representada em termos de axiomas universalmente válidos. Para lidar com tais variedade e novidade, precisamos de estilos de pensamento mais flexíveis, tais como os que vemos no raciocínio humano do senso comum, o qual é baseado mais em analogias e aproximações do que em procedimentos formais precisos<sup>20</sup>.

Para Minsky, a questão relevante em relação à IA é como modelar computacionalmente ações que estão atreladas ao senso comum, ou seja, ao cotidiano da vida. Assim, Minsky aponta para o fato de que os pesquisadores da IA, tradicionalmente, almejam modelar ações lógicas que necessitam de reflexões prévias e o uso de regras (cálculos matemáticos, por exemplo) e minimizam as ações cotidianas que os seres humanos normalmente e naturalmente realizam (identificar contextos e expressões faciais, por exemplo). Tal minimização pode ser justificada pela indisponibilidade de tecnologias de programação da época, mas como apontamos o fato de, tradicionalmente, ações puramente lógicas serem consideradas como inteligentes em detrimento de ações práticas revela um tipo de pensamento vigente entre os pesquisadores da área.

Todavia, para Minsky (1991, p. 36), as técnicas *top-down* são satisfatórias no que diz respeito a realizar procedimentos de buscas eficientes, manipulação e reorganização de elementos de situações complexas. Tais ações desempenhadas eficientemente pelas técnicas *top-down* são um tipo de ações que os seres humanos efetuam. E dada a falha das técnicas *top-down* para a realização de outros tipos de ações (as do senso comum, por exemplo),

<sup>20</sup>[...] the objects and activities of everyday life are too endlessly varied to be described by precise, logical definitions and deductions. Commonsense reality is too disorderly to represent in terms of universally valid axioms. To deal with such variety and novelty, we need more flexible styles of thought, such as those we see in human commonsense reasoning, which is based more on analogies and approximations than on precise formal procedures.

Minsky aponta para outro paradigma de modelagem: a abordagem *bottom-up*. Tal abordagem é caracterizada por iniciar o processo de modelagem a partir de elementos simplificados e ascender gradualmente em uma escala de complexidade de modo a interconectar os elementos simplificados para produzir um fenômeno emergente de grande escala.

Para Minsky, tanto as técnicas *top-down* quanto as *bottom-up* possuem limitações na modelagem de comportamentos inteligentes. Nas palavras de Minsky:

Os sistemas top-down são prejudicados por mecanismos inflexíveis para lidar com interações muito numerosas, embora muito fracas, enquanto sistemas bottom-up são limitados por arquiteturas inflexíveis e limitações organizacionais. Nenhum tipo de sistema foi desenvolvido para ser capaz de explorar variedades de conhecimento múltiplas e diversas (MINSKY, 1991, p. 37, tradução nossa)<sup>21</sup>.

A questão que Minsky coloca é: Porque tanto a abordagem *top-down* quanto a *bottom-up* foram desenvolvidas de modo a ser tornarem tão inflexíveis? Cada abordagem possui vantagens e desvantagens, então, para a modelagem satisfatória é necessário o desenvolvimento de sistemas interligados capazes de explorar as vantagens de ambas. Para Minsky não se trata apenas de questionar como as máquinas ‘pensam’, mas como as pessoas pensam. Minsky aponta que: “[...] não há uma maneira mágica e simples de evitar tais problemas em nossas novas máquinas; isso vai requerer um grande esforço de pesquisa e de engenharia<sup>22</sup>” (MINSKY, 1991, p. 51, tradução nossa).

Antes mesmo das críticas elaboradas por cientistas cognitivos como Marvin Minsky, filósofos como Searle e Dreyfus problematizaram o projeto da IA. John Searle em seu artigo *Minds, Brains and Programs* critica os princípios basilares da IA. Inicialmente o filósofo questiona: “Qual significado psicológico e filosófico devemos atribuir aos esforços recentes dirigidos à produção de simulações computacionais de capacidades cognitivas humanas?<sup>23</sup>” (SEARLE, 1980, p. 418, tradução nossa). Em resposta a essa questão, Searle distingue o ramo da IA em duas vertentes com modelagens distintas: a IA forte (*strong AI*) e a IA fraca (*weak AI*). No sentido “fraco” o computador é apenas uma mera ferramenta de simulação no estudo da mente. Por outro lado, segundo a IA forte, o computador não é

---

<sup>21</sup> The top-down systems are handicapped by inflexible mechanisms for dealing with very numerous, albeit very weak, interactions, while the bottom-up systems are crippled by inflexible architectures and organizational limitations. Neither type of system has been developed to be able to exploit multiple, diverse varieties of knowledge.

<sup>22</sup> I suspect there is no simple and magical way to avoid such problems in our new machines; it will require a great deal of research and engineering.

<sup>23</sup> What psychological and philosophical significance should we attach to recent efforts at computer simulations of human cognitive capacities?

meramente um meio para o estudo da mente, ele pode ser a mente em si, no sentido que se pode dizer literalmente que os computadores, munidos de programas certos, compreendem e têm estados mentais. Na IA forte, como o computador possuiria estados mentais, os programas não são meramente ferramentas que permitem testar explicações psicológicas; pelo contrário, os programas são em si as próprias explicações.

Searle também focalizou a falha semântica dos modelos da IA tradicional, ou seja, para a incapacidade de tais modelos capturarem significados em um contexto. Para fundamentar sua crítica, Searle se utiliza de um experimento mental que consiste em se imaginar preso em um quarto. A pessoa presa no quarto apenas domina e sabe como se comunicar segundo sua língua nativa. Dentro do quarto existe um livro de instruções, escrito na língua materna da pessoa, que permite manipular sintaticamente símbolos chineses. A pessoa recebe periodicamente textos e tarefas escritas em chinês e deve, segundo o livro de instruções, realizar o que se pede. A pessoa mesmo não sabendo chinês consegue, com certo êxito, realizar o que se pede nos textos recebidos em chinês. Para Searle, a pessoa *simula* a capacidade de compreender chinês, mas não possui a efetiva capacidade de entender o idioma chinês. Através do célebre “Argumento do quarto chinês”, Searle ilustra a incapacidade semântica que, segundo ele, caracteriza os modelos da IA, pois eles consistem apenas em manipulações de símbolos a partir de um conjunto de regras sem abarcar o significado. Nas palavras de Searle (1980, p. 427, tradução nossa) “[...] a manipulação de símbolos formais, por si só, não tem intencionalidade, não é sequer manipulação de símbolos, uma vez que estes símbolos não simbolizam nada. No jargão da linguística, eles têm apenas sintaxe, mas não semântica<sup>24</sup>”. Os modelos computacionais apresentam um comportamento específico para uma dada situação, porém este comportamento é destituído de significado. Searle defende a hipótese de que, no que diz respeito à ontologia da mente, o comportamento pode ser considerado de menor relevância (SEARLE, 1997, p. 115). Nesse caso, os modelos da IA satisfariam apenas os quesitos de simulação e não abarcariam de maneira satisfatória a realidade biológica da mente.

Para Searle (1997, p. 7) “os fenômenos mentais são causados por processos neurofisiológicos no cérebro e, são, eles próprios, características do cérebro.”. Tais processos não poderiam ser instanciados em sistemas artificiais por uma limitação técnica. Ou seja, certas capacidades são exclusivamente humanas, dado seu caráter biológico. Searle dá como exemplo o modo como reagimos a expressões faciais, ou seja, a maneira pela qual inferimos

---

<sup>24</sup> [...] the formal symbol manipulations by themselves don't have any intentionality; they are quite meaningless; they aren't even symbol manipulations, since the symbols don't symbolize anything. In the linguistic jargon, they have only a syntax but no semantics.

pela expressão pública do rosto de uma pessoa seu estado emocional, algo considerado natural no comportamento.

No modo como interpretamos as feições, em grande parte, estão envolvidas ferramentas derivadas do *background* da pessoa que interpreta, que são derivadas por sua vez, da experiência vivenciada pela pessoa no plano da ação. Quando vemos uma pessoa chorar, inferimos, mesmo sem ter algum contato mais próximo com essa pessoa, que ela está triste. Perceber a tristeza a partir das lágrimas é uma atividade referente a uma experiência passada que constitui nosso *background*: já vimos, outrora, outra pessoa chorar e esta nos falou que chorava porque estava triste com a morte de um amigo, por exemplo. A experiência que tivemos e a experiência que temos agora não estão no campo proposicional, não são abarcadas por uma teoria, mas são de natureza prática e estão no campo da experiência. É sabido por, exemplo, que um gol em uma partida de futebol tem valor de 1 (um) ponto. Se alguém formular uma teoria e dizer que um gol tem valor, na verdade, de 0.99999 pontos, mesmo que essa teoria seja assentada em cálculos complexos fundamentados nas leis da física, normalmente aqueles que entendem do esporte dirão que esta pessoa está muito confusa, pois é correntemente considerado que um gol é um ponto em uma partida de futebol. Teorizar sobre isso pode até alterar a definição, mas não descobrir um fato diferente. A “crença” de que um gol vale um ponto está no campo da experiência do indivíduo e, por isso, é válida por si só, mesmo que não seja passível de comprovação empírica (em seu sentido científico e estrito). Nesse sentido, a crítica que Searle tece à IA simbólica se refere a que os modelos assim construídos são limitados ao campo proposicional, no qual o conhecimento comum da vida cotidiana não está abarcado.

Como apontado por Haselager (2004, p. 107), “Um dos mais importantes problemas enfrentados pela Ciência Cognitiva é como fazer com que computadores apresentem senso comum e executem tarefas do cotidiano.”. A dificuldade de modelar tais tarefas diz respeito à presença intrínseca de conhecimento comum em seu exercício. Tradicionalmente, as tentativas de modelar o conhecimento comum se assentaram em sistemas internos representacionais, todavia, sistemas com uma grande gama de representações tendem a fracassar, pois o sistema acaba se perdendo em seu próprio armazenamento de informações. Essa problemática traz à baila o “problema do *frame*”, ou seja, o problema de como representar satisfatoriamente o conhecimento cotidiano e de distinguir fatores mais ou menos relevantes em contextos dados. Tal problema surge da tentativa de formalizar o processo de resolução de problemas envolvendo o mundo dinâmico e complexo. Teixeira (1998, p. 63) caracteriza *frame* como “uma estrutura hierárquica de nós e relações onde os nós superiores

representam conceitos gerais e os inferiores instâncias específicas desses conceitos”. Para elucidar tal conceito Minsky, em *Society of mind* apresenta o seguinte exemplo:

“Mary foi convidada para a festa de Jack.

Ficou imaginando se ele gostaria de ganhar uma pipa” (MINSKY, 1986, p. 261)<sup>25</sup>

Como apontado por Teixeira (1998, p. 64), a sentença “Mary foi convidada para a festa de Jack” ativa o *frame* “convite para festa”, que consequentemente ativa *frames* subsequentes do tipo “quem é o anfitrião?”, “Que roupa usar?”, “Que presente levar?”. E cada um desses *frames*, por sua vez, está conectado a um outro frame e assim consequentemente, formando uma *rede* complexa de conhecimentos interdependentes. Em suma, o problema do *frame* decorre da dificuldade de estabelecer um critério delimitador do conhecimento relevante à um contexto dado.

Nesse mesmo viés, o filósofo Hubert Dreyfus (1975, 1993) também se coloca como crítico da IA tradicional. Dreyfus questiona princípios basilares da IA e a maneira como esta concebe o comportamento inteligente que pretende modelar. Dreyfus indaga:

(1) No ‘processamento de informações’ um ser humano obedece, de fato, a regras formais como se fosse um computador digital? e (2) Pode o comportamento humano, não importa qual seja a sua gênese, ser descrito com tal formalismo que possa ser manipulado por uma máquina digital? (DREYFUS, 1975, p. 197)

O filósofo defende a hipótese de que há capacidades não modeláveis presentes em todas as formas de comportamento inteligente. Segundo Dreyfus (1975), os argumentos dos pesquisadores da IA tradicional parecem se basear justamente nas hipóteses de que o mundo pode ser formalmente decomposto em elementos lógicos e de que a compreensão acerca do mundo pode ser reconstruída na combinação de elementos por meio de regras. A hipótese de que o comportamento deve ser compreendido como sendo guiado por um conjunto de instruções pré-determinadas implica a regressão às regras para aplicação de regras. Dreyfus aponta para que em situações cotidianas, como a participação em um jogo, a resolução de problemas e o reconhecimento de feições<sup>26</sup>, por exemplo, parece que os seres humanos não

---

<sup>25</sup> Mary was invited to lack's party.  
She wondered if he would like a kite.

<sup>26</sup> Esta crítica de Dreyfus é bem anterior aos atuais modelos de reconhecimento de expressões faciais como os propostos por Du, Tao e Martinez (2014) que são capazes de reconhecer vinte e uma diferentes expressões faciais.

estão seguindo regras, mas utilizando uma organização perceptual global e o conhecimento do senso comum referente à situação.

Além disso, Dreyfus argumenta que os elementos que devem ser reordenados seguindo pressupostos lógicos, para adquirirem formas significativas (como defende a IA tradicional) estão sujeitos a diversas interpretações. Será o contexto que determinará qual regra deverá ser aplicada ou se uma nova deverá ser produzida. Segundo Dreyfus (1975), para um computador, entretanto, o próprio contexto apenas pode ser reconhecido de acordo com uma regra prévia.

Nesse contexto, segundo Dreyfus (1975, p. 147), a IA tradicional falha ao desconsiderar o papel exercido pelo corpo nos comportamentos inteligentes. Mas se considerarmos o corpo como algo indispensável ao comportamento inteligente, teremos que refletir sobre a possibilidade de modelá-lo em um computador digital programável (DREYFUS, 1975, p. 209). Para Dreyfus, o corpo tem a capacidade de modificar constantemente suas ações sem necessariamente teorizar ou produzir hipóteses sobre a realidade. O corpo se adapta aos desafios do mundo porque ele está completamente situado nele. Nas palavras do filósofo:

Os partidários das hipóteses psicológicas e epistemológicas de que o comportamento humano deve ser formalizado em função de um programa heurístico para computador digital são forçados a desenvolver uma teoria do comportamento inteligente que despreza o fato do ser humano ser dotado de um corpo, já que, pelo menos até agora, é evidente que os computadores não o têm (DREYFUS, 1975, p. 209).

A ciência da computação tem atingido certo êxito técnico na *simulação* das funções racionais: os programas de computadores lidam satisfatoriamente com linguagens ideais e relações lógicas abstratas, mas, segundo Dreyfus (1975, p. 211), deixam a desejar na simulação do tipo de inteligência que compartilhamos com os animais, como o reconhecimento de padrões. Tal inteligência diz respeito à capacidade natural de antecipação indeterminada e global. Por exemplo, quando reconhecemos uma melodia, as notas adquirem sentido ao serem percebidas como parte da melodia, como um todo, e não individualmente. Nosso senso natural de totalidade situacional e as experiências passadas nos dão uma percepção do conjunto e nos orientam no preenchimento dos detalhes (DREYFUS, 1975, p. 217). Para o filósofo, os computadores não possuem e nem podem possuir tal capacidade, pois trata-se de uma limitação lógica e biológica da modelagem computacional. Um programa capaz de manipular uma base de dados pré-determinada de acordo com regras definidas poderia verificar dentro de um conjunto fixo de possibilidades qual seria a melhor hipótese

para explicar uma dada situação, mas, como aponta Dreyfus (1975, p. 218): “[...] isso está longe de corresponder à interação flexível de dados indeterminados e expectativas indeterminadas que parecem ser características do reconhecimento humano de padrões”.

O que Dreyfus pretende defender é que o comportamento humano considerado inteligente pode ser regular sem ser, necessariamente, normativo. Em outras palavras, a inteligência humana pode não ser regida exclusivamente por regras formalizáveis e pode não ser um sistema global de regras cuja aplicação a todas as eventualidades possíveis seja antecipadamente determinada. Como um computador não está situado no contexto (o próprio contexto deve ser *programado* para a máquina), todos os fatos tornam-se possivelmente relevantes em todas as situações. As máquinas não possuem critério de relevância e para Dreyfus tal habilidade não é modelável computacionalmente, pois nem mesmo nós sabemos definir como ocorre tal processo em nós mesmos. O ser humano está naturalmente situado em seu mundo e isso lhe permite operar sobre ele. Nas palavras de Dreyfus:

Os seres humanos estão situados no mundo de coisas que o cercam de tal maneira, que aquilo de que necessitam para lidar com as coisas não está embrulhado longe deles dentro de uma arca, e sim situado a seu alcance, nos lugares mais convenientes, ou então jaz cuidadosamente como se fosse uma coleção de fichas num arquivo. O sistema de relações que torna possível descobrir objetos quando estes se tornam necessários é o nosso lar, ou nosso mundo. (DREYFUS, 1975, p. 239)

Desse modo, Dreyfus aponta para o fato de estarmos situados e familiarizados com o mundo e, assim, podermos nos orientar nele pelo simples fato dele ser o *nosso* mundo, ao qual nos adaptamos em processos histórico biológicos de longa duração e que foi, em um certo sentido, produzido por nós através de nossas atividades. O mundo estaria estruturado de acordo com as tarefas dos seres que nele habitam naturalmente. Para Dreyfus nós somos seres que modelam o seu próprio mundo e tal fato não pode ser representado mecanicamente, de modo que uma máquina seja capaz de modelar o próprio mundo.

Em consonância com as críticas tecidas por Dreyfus à IA tradicional, filósofos como Andy Clark (1999) e Antony Chemero (2007) criticam o representacionismo presente nos modelos computacionais, na medida em que propõem a análise das habilidades cognitivas como resultado das relações estabelecidas entre o agente e o meio. Tais críticas constituem o escopo da Teoria da Cognição Incorporada e Situada e que abordaremos no próximo Capítulo.

## **CAPÍTULO 2- A ABORDAGEM EXTERNALISTA DA MENTE E A MODELAGEM COMPUTACIONAL**

*[...] the problem of rationality becomes, precisely, the problem of explaining the production, in social, environmental and emotional context, of broadly appropriate adaptive response. Rationality (or as much of it as we humans typically enjoy) is what you get when this whole medley of factors are tuned and interanimated in a certain way. Figuring out this complex ecological balancing act just is figuring out how rationality is mechanically possible (Andy Clark, 2003, p. 18).*

## APRESENTAÇÃO

Neste Capítulo, a partir da análise das críticas dirigidas ao paradigma de modelagem tradicional, apontaremos para um outro tipo de abordagem no que fiz respeito à análise dos comportamentos inteligentes. Sendo assim, desenvolvemos o presente Capítulo, como se segue: na **seção 2.1**, apresentaremos os pressupostos da Teoria da Cognição Incorporada e Situada, enquanto uma vertente que considera o indivíduo cognoscente enquanto um agente que interage com o meio. Na **seção 2.2**, relacionaremos as teses da Teoria da Cognição Incorporada e Situada com o pensamento de Gilbert Ryle enquanto crítico do representacionismo. Na **seção 2.3**, analisaremos os conceitos de *saber como* e *saber que* presentes na obra de Ryle, a fim de sublinhar a hipótese segundo a qual a cognição é um processo de interação entre um agente e o meio no qual ele está inserido. Nesse sentido, exporemos o conceito de *agente disposicional* enquanto uma alternativa de modelagem. Por fim, na **seção 2.4** apresentaremos a Teoria da Auto-organização proposta por Debrun, relacionando-a com o desenvolvimento de habilidades decorrente das disposições de um agente situado e incorporado.

### 2.1 A TEORIA DA COGNIÇÃO INCORPORADA E SITUADA (CIS)

Na Ciência Cognitiva tradicional é considerado que são os estados mentais, enquanto representações da realidade, que causam o comportamento inteligente. Isso desencadearia a interpretação de que crenças e desejos são basicamente proposicionais, ou seja, que o conhecimento comum é descritivo, pois *descreve* uma parte da realidade apreendida. Desse modo, admite-se que somos capazes de agir no mundo uma vez que possuímos representações deste e, como tais representações são internas, a cognição é interpretada conforme o modelo de mundo a que temos acesso e segundo o qual raciocinamos e planejamos antes de agir (HASELAGER, 2004, p. 215-216). Em outras palavras, a hipótese da Ciência Cognitiva tradicional é de que os seres vivos representam estímulos recebidos do meio, criam representações internas do ambiente, consultam suas crenças e desejos, geram planejamentos e por fim elegem um planejamento para ser executado a fim de produzir um comportamento adequado. Tal postura, enquanto opção metodológica, não considera que também nos comportamos mecanicamente e “racionaliza” o comportamento comum, considerando este tipo de comportamento também como um processo de manipulação de símbolos.

A cognição, segundo a Ciência Cognitiva tradicional, se equipara ao processo de manipulação de símbolos de uma calculadora. Uma máquina de calcular pode produzir o símbolo ‘5’, em resposta à entrada ‘2’, ‘+’, ‘3’, estes símbolos não possuem significado para a

máquina, mas para o usuário (aquele que inseriu a entrada) o resultado ‘5’ “faz sentido” dado o significado que este usuário atribui aos símbolos (SHAPIRO, 2007, p. 339). Tal visão aponta para a abordagem *computacional* da mente, cujas teses centrais apresentamos na primeira parte deste trabalho, segundo a qual a mente de um ser vivo pode desempenhar a mesma função de um *hardware* de computador (capta estímulos do meio e as representa de acordo com regras inatas ou aprendidas e, por fim, desencadeia uma dada ação). A mente, nesse contexto, seria programas executados em *hardwares* e “... a cognição é separada do resto do mundo no sentido de que os processos cognitivos operam na liberdade simbólica dos órgãos dos sentidos<sup>27</sup>” (SHAPIRO, 2007, p. 339, tradução nossa).

Como vimos, a mais relevante crítica que se faz ao modelo da cognição da Ciência Cognitiva tradicional diz respeito, justamente, aos projetos de sucesso serem baseados na manipulação de símbolos e, assim, facilmente descritos através de algoritmos (como jogar xadrez). O limite da modelagem se dá na tentativa de construir um modelo capaz de se movimentar em um ambiente desordenado não controlado, por exemplo. Esse tipo de atividade parece requerer capacidades muito mais simples do que a de jogar xadrez e, mesmo assim, ainda se apresenta como um desafio para a Ciência Cognitiva tradicional. Em resposta à estas dificuldades, os pesquisadores da teoria da cognição situada e incorporada minimizam o papel das representações no processo de cognição. Como apontado por Shapiro (2007, p. 340, tradução nossa):

A ideia de que um organismo deve consultar uma representação convida um tipo vicioso de homuncularismo. A representação do mundo também deve ser representada para que ela seja útil? Se assim for, uma regressão ao infinito começa. Se não, se a representação do mundo pode guiar a ação em si, sem ser representada, porque o resto do mundo não o pode fazer?<sup>28</sup>

Uma hipótese central que a teoria da cognição situada e incorporada defende é, além de considerar o ambiente na cognição, maximiza o papel do corpo na explicação das habilidades cognitivas. Segundo esta teoria, o processo de cognição pode emergir da interação dos atributos físicos dos corpos com o meio ambiente em que estes corpos estão situados. Para Clark: “A cognição não é um fenômeno que pode ser estudado com sucesso enquanto se

---

<sup>27</sup>Cognition is cut off from the world in the sense that cognitive processes operate only on symbolic deliverances from the sense organs.

<sup>28</sup>The idea that an organism must consult a representation invites a vicious sort of homuncularism. Must the representation of the world also be represented if it is to be useful? If so, an infinite regress begins. If not, if the representation of the world can guide action without itself being represented, why can't the world do the same?

marginaliza os papéis do corpo, do mundo e da ação<sup>29</sup>” (CLARK, 1999, p. 350, tradução nossa). Para Clark, as capacidades cognitivas humanas dependem profundamente do entorno em que seus agentes epistêmicos estão inseridos. Todavia, o filósofo defende um *externalismo ativo*, destoando, assim, da proposta externalista de Putnam (1975) e Burge (1979). Para Clark, o entorno exerce um poder causal no processo de conhecimento, e assim “Se eliminarmos o componente externo do sistema de condutas, é como se eliminássemos uma parte do próprio cérebro<sup>30</sup>” (CLARK, 1998, p. 13, tradução nossa). A tese de Clark diz respeito à hipótese de que os processos cognitivos podem (ou não) ocorrer *dentro do cérebro*. Tal tese aponta para a hipótese de que os fatores externos relevantes são ativos e que tem um papel crucial no momento presente e, dada a sua complementariedade com o organismo humano, acabam tendo um impacto direto no organismo e em sua conduta. Trata-se de estar *situado* em um ambiente e *incorporar* informações relevantes para um dado agente, que é ativo e não apenas representativo. Ao defender o externalismo ativo, Clark traz à baila uma explicação mais natural das ações e do comportamento inteligente.

Chemero (2007), por sua vez, considera que, uma vez os agentes situados em um ambiente, o papel da representação mental no desenvolvimento de habilidades é minimizado, pois “[...] uma parte da inteligência é ‘carregada’ do cérebro para o corpo e o ambiente<sup>31</sup>” (CHEMERO, 2007, p. 181, tradução nossa). Nessa visão, nosso corpo constitui uma ferramenta adequada cujo *design* a torna fácil de ser manipulada pelo cérebro. Por exemplo: nossos joelhos limitam os movimentos de nossas pernas, mas tornam o equilíbrio corporal e a locomoção possíveis. É através do corpo que conhecemos o ambiente e reconhecemos neste as possibilidades de nossas ações (CHEMERO, 2009, p. 27). O corpo é importante, na abordagem de Chemero, pois é através dele que a mente pode fazer algo: o corpo, então, pode influenciar os processos mentais. Nessa perspectiva, o processo de aquisição e desenvolvimento de habilidades não é reduzido à manipulação de representações mentais. Os adeptos da Teoria da Cognição Incorporada e Situada defendem a hipótese segundo a qual existe uma dinâmica intrínseca entre os seres e seus ambientes (HASELAGER, 2004, p.221).

Os pressupostos da cognição situada e incorporada apontam para a hipótese que nossas ações não são necessárias e exclusivamente direcionadas pelas representações internas cerebrais, mas dependem também de processos dinâmicos estabelecidos entre o corpo e o

---

<sup>29</sup>Cognition is not a phenomenon that can be successfully studied while marginalizing the roles of body, world and action.

<sup>30</sup> If we remove the external component the system's behavioral competence will drop, just as it would if we removed part of its brain.

<sup>31</sup>[...] some of the intelligence is “off-loaded” from the brain to the body and environment.

ambiente. Tal processo é de natureza fluída, contínua e recorrente (HASELAGER, 2004, p. 224).

Um exemplo que ressalta o papel do meio ambiente para a cognição foi apontado por Kirsh e Maglio (1994). Através da análise do jogo de computador “Tetris”, os autores apontam para a existência de dois tipos de ações para a resolução dos problemas de encaixe entre as peças. A primeira seria mover mentalmente as peças, representando um encaixe e o segundo tipo de ação seria a rotação efetiva da peça através do console do jogo. Testando essas ações, Kirsh e Maglio chegam à conclusão de que a rotação mental demora 1000ms enquanto a rotação “prática” realizada fisicamente através do console do computador demora 400 ms (KIRSH; MAGLIO, 1994, p. 533). Para os autores, a ação de usar um controle remoto para jogar um jogo é uma ação *epistêmica* que se revela eficaz no plano da ação.

Nesse sentido, a problemática em relação à IA contemporânea consiste justamente na possibilidade de modelos computacionais estarem situados no meio ambiente e serem realmente capazes de adquirir e atualizar seus estados e funções de acordo com esse meio. Esse tipo de modelagem de inspiração externalista contemplaria o objetivo da IA ao modelar o comportamento inteligente. No presente trabalho, sugerimos que a abordagem de Gilbert Ryle (a qual propõe que há ações inteligentes incorporadas nos agentes que sabem como desempenhar uma tarefa com habilidade) e as teses da Cognição Incorporada e Situada estariam no cerne de uma nova abordagem dos modelos computacionais. Explicitaremos as teses centrais propostas por Gilbert Ryle na sessão a seguir.

## 2.2 RYLE CRÍTICO DO REPRESENTACIONISMO

Ryle se apresenta como crítico dos pressupostos internalistas, principalmente no que diz respeito ao aspecto de “dupla vida” do sujeito cartesiano. Nesse sentido, segundo a tradição internalista dualista substancial, todo indivíduo tem um corpo e uma mente distintos, mas unidos momentaneamente. Em consequência, todo indivíduo viveria duas histórias paralelas: uma seria constituída pelos eventos acontecidos ao corpo e a outra, pelos eventos mentais. A história vivida pelo corpo é pública e acessível a observadores, pois os eventos constituintes desta história pertencem ao mundo físico. Já a história mental do indivíduo seria acessível a ele próprio, de forma direta por meio da consciência, autoconsciência e introspecção (RYLE, 1949, p. 16).

Podemos dizer que a abordagem dualista substancial acerca da relação mente-corpo, considerada hegemônica por Ryle, apresenta uma contraposição entre estes termos. Tal abordagem divide a existência humana em duas vidas e em dois mundos distintos: o corpo seria externo e as operações da mente, internas. Esta antítese entre interno e externo, no entanto, parece existir metaforicamente, uma vez que a mente não é espacial e por isso não pode estar ou ser colocada em lugar algum (nem dentro, nem fora do próprio corpo). Segundo Ryle, apenas a metáfora salva a doutrina dualista substancial da contradição. Todavia, as conexões efetivas entre os episódios da vida privada (da mente) e os da vida pública (do corpo) permanecem mal explicadas, pois, segundo Ryle, estas conexões não pertencem ao mundo físico nem ao mundo mental.

Desse modo, Ryle considera que os princípios centrais da doutrina internalista, chamada por ele de *doutrina oficial*, se apoiam em pressupostos equivocados (RYLE, 1949, p. 10). Segundo Ryle a tradição internalista comete um tipo de equívoco lógico. Nas palavras do filósofo, tal equívoco é do tipo categorial e “(...) apresenta os fatos da vida mental como se pertencessem a um tipo de categoria lógica (ou classes de tipos ou categorias) quando efetivamente eles pertencem a outra categoria” (RYLE, 1949, p. 17, tradução nossa)<sup>32</sup>.

Para ilustrar o “erro categorial”, Ryle apresenta o exemplo da universidade. A um visitante que, pela primeira vez conhece uma universidade, lhe são mostradas as salas de aulas, bibliotecas, campos de esportes, departamentos e a administração. Mas, o visitante pergunta “Onde está a universidade? Vi onde as aulas são ministradas e onde se faz os experimentos científicos, mas não vi a universidade”. Ryle observa que o visitante, ao perguntar pela universidade, espera receber uma resposta como: “a Universidade está ao lado da sala de aula”, e comete um erro categorial. A universidade é exatamente o conceito funcional que abarca o conjunto das salas de aulas, bibliotecas, campos de esportes, pessoas, departamentos e administração. O visitante se equivoca ao considerar que “a universidade” se refere a um membro adicional do conjunto do qual fazem parte os outros elementos constituintes do conceito “universidade”. Erroneamente ele considerou a universidade como pertencente à mesma categoria à qual pertencem os outros elementos constituintes. Ou seja, a vertente internalista de inspiração cartesiana apresenta mente e corpo como pertencendo à categoria lógica **substância**. Na abordagem internalista cartesiana, como vimos, o sujeito é uma “substância pensante” de natureza distinta do corpo e que conhece apenas à medida que elementos constituintes de um conhecimento possível passam pelas representações da razão.

---

<sup>32</sup> It represents the facts of mental life as if they belonged to one logical type or category (or range of types or categories), when they actually belong to another.

Segundo Ryle, o erro categorial característicos da doutrina internalista cartesiana tem origem quando Galileu mostrou que seu método de investigação científica era apto para proporcionar uma teoria mecânica aplicável a todo corpo físico. Nesse momento, Descartes se encontrou em uma situação conflitante, pois como cientista não podia deixar de apoiar as pretensões da mecânica, mas como homem religioso e moral não podia aceitar a consequência de que a natureza humana fosse essencialmente mecânica: recusava-se a considerar que a mente pudesse ser uma mera variedade do mecânico.

Como o vocabulário mecanicista de então não parecia adequado para tratar da complexidade da mente, Descartes e outros filósofos optaram por entender o mental significando o acontecimento de processos não mecânicos. No entanto, permanece na teoria da mente o molde lógico que Galileu usou em sua teoria mecânica: Descartes apropriou-se de um modelo em linguagem mecânica e o adequou para descrever a mente, transcrevendo-o em um vocabulário puramente negativo (a mente seria i-material, in-extensa, não ocuparia lugar no espaço e não estaria sujeita às leis mecânicas). Para Ryle, enquanto representada como negação das propriedades mecânicas, a mente se mostra como um **fantasma** dentro do corpo humano, e nada se sabe de seu comportamento. A imagem do “fantasma da máquina” se cristaliza na combinação da suposição de que teorizar é a atividade primária da mente e que é, necessariamente, uma atividade privada e interna.

Embora Ryle não se proponha de fato a elaborar uma teoria da mente, sua concepção acerca das teorias internalistas, principalmente a cartesiana, aponta para a necessidade de uma reformulação da noção clássica de sujeito enquanto entidade metafísica. Tal sujeito, no escopo do pensamento de Ryle, é substituído por um agente que atualiza suas disposições pela interação com o meio ambiente através do conjunto de suas ações.

Ryle observa que a doutrina cartesiana pressupõe uma concepção de mente responsável pela representação isolada da realidade através de processos internos. Esta concepção considera que o conhecimento praticamente se reduz à capacidade de teorizar, gerar abstrações e de realizar outras atividades intelectuais consideradas de alto nível, já que a interação perceptiva não forneceria informações confiáveis sobre os objetos e os fatos do mundo. A ação, então, seria guiada pela mente como resultado de deliberações prévias que a antecederiam necessariamente. Analisaremos, a seguir o conceito de ‘saber que’ privilegiado pela doutrina internalista e o enriquecido com o conceito de ‘saber como’, conforme proposto por Gilbert Ryle (1949).

A partir dos questionamentos atuais em contraste à tradição internalista, os agentes deixam de ser considerados seres ontologicamente fragmentados e parcialmente distintos da

realidade externa. Nesse sentido, as capacidades cognitivas deixam de ser consideradas internas e particulares, mas passam a ser concebidas como resultado de uma constante dinâmica entre as estruturas de um agente e o meio em que está situado. Em seu livro *The concept of mind*, Ryle destaca que o modo de se obter conhecimento no viés internalista pode ser descrito pelo termo “*saber que*”. Este, segundo Ryle (1949, p. 29), pressupõe necessariamente a precedência e o domínio de um aparato de regras e teorias anterior a, e causa de, uma atividade inteligente. Para Ryle (1949, p. 26, tradução nossa) “A prática inteligente não é um desmembramento da teoria. Pelo contrário, teorizar é uma prática entre outras, que pode ser executada com inteligência ou estupidez”<sup>33</sup>. Ser inteligente, nesse sentido, não consiste apenas em satisfazer critérios, mas em saber aplicá-los.

Na perspectiva de Ryle, o conceito de “*saber que*” por si só não é capaz de explicar muitos fenômenos cognitivos referentes ao desenvolvimento de habilidades práticas que podem ser reconhecidas como ações inteligentes. Este tipo de habilidade de um agente ambientalmente situado não pode ser reduzido a capacidades lógico-inferenciais. Nesse viés, o “*saber como*” se efetiva através de uma constante atualização de potencialidades disposicionais da ação, Ryle apresenta o exemplo de um estudante estrangeiro que pode não saber como falar corretamente o idioma do país que visita, tão bem quanto um indivíduo nativo desta língua. O nativo sabe como falar e a estrutura da língua que lhe é natural, pois é apreensível em seu meio ambiente desde seu nascimento. Uma criança francesa não representa o francês e depois fala o idioma, ela aprende a falar francês “falando”, exercitando-se constantemente, criando e fortalecendo hábitos ao repetir inteligentemente comportamentos bem sucedidos. Em outras palavras, em uma parte muito significativa do exercício de habilidades não cabe a clássica dissociação teoria-prática. Trata-se de um conhecimento pragmático, de um “*saber como*”, expresso em ações **cognitivamente carregadas** praticadas por um agente (e não mais por um sujeito concebido como pensamento puro).

O conceito de sujeito clássico e metafísico, no escopo do pensamento de Ryle, é substituído pelo conceito de *agente* contextualmente situado capaz de atualizar habilidades pela aprendizagem. Habilidades desse tipo, dado seu caráter inteligente, não podem ser reduzidas a hábitos, segundo Ryle, uma vez que estes se caracterizam por resultarem de réplicas mecânicas de ações anteriores, aprendidas por repetição e condicionamento. Para Ryle “Quando dizemos que alguém faz algo por puro hábito, queremos dizer que o faz automaticamente, sem ter consciência do que está fazendo” (RYLE, 1949, p. 42, tradução

---

<sup>33</sup>Intelligent practice is not a step-child of theory. On the contrary theorizing is one practice amongst others and is itself intelligently or stupidly conducted.

nossa)<sup>34</sup>. As habilidades, por sua vez, embora também dependam de repetições, são adquiridas pela aprendizagem, de modo que cada ação executada com cuidado e atenção se torna uma nova ferramenta cognitiva que contribui para seu aprimoramento. O “saber como”, não é uma repetição simples como um hábito, mas carrega em si uma *disposição* para aperfeiçoar-se que um agente possui.

### 2.3 O AGENTE DISPOSICIONAL E O NEODISPOSICIONALISMO

Segundo Ryle (1949), a ação inteligente não necessitaria somente de um conhecimento intelectual, pois não haveria divisão efetiva entre pensar e agir: “aprendemos a agir mediante a prática, ajudados pela crítica e pelo exemplo, embora frequentemente sem receber lição alguma sobre a teoria.”<sup>35</sup>. (1949, p. 41, tradução nossa). Aprendemos a atuar com habilidade no mundo mediante um constante aperfeiçoamento de habilidades adquiridas, atualizando disposições e propiciando a emergência de um novo conjunto disposicional que possibilita a incorporação de novas habilidades. Ryle (1949, p. 112) propõe que certas atividades cognitivas complexas podem ser explicadas pela noção de *disposição*. Tal noção deriva como resultado de uma série de termos disposicionais apresentados por Ryle e que dizem respeito à detecção e atualização de tendências de conduta.

Como apontado por Ryle: “Termos como ‘conhecer’, ‘acreditar’, ‘esperar’, ‘inteligente’ e ‘humorado’ são termos disposicionais determináveis. Significam habilidades, tendências ou propensões a fazer algo que não pertencem a um único tipo, mas que pertencem a muitos tipos diferentes” (RYLE, 1949, p. 118)<sup>36</sup>. Desse modo, o conceito de disposição estaria na gênese das explicações de fatos ou eventos no mundo, uma vez que possibilita a compreensão da atualização de uma potencialidade natural possibilitada pelo meio. Ou seja, a disposição de um agente mostra as possibilidades deste em contato com o mundo, podendo ser uma ferramenta explicativa da conduta. Para Ryle:

Quando dizemos que o vidro é frágil ou que o açúcar é solúvel, usamos conceitos disposicionais cuja força lógica é a seguinte. A fragilidade do vidro não consiste no fato de que, em um dado momento, se faça em

---

<sup>34</sup> When we describe someone as doing something by pure or blind habit, we mean that He does it automatically and without having to mind what he is doing.

<sup>35</sup> We learn *how* by practice, schooled indeed by criticism and example, but often quite unaided by any lessons in the theory.

<sup>36</sup> Dispositional words like ‘know’, ‘believe’, ‘aspire’, ‘clever’ and ‘humorous’ are determinable dispositional words. They signify abilities, tendencies or proneness to do, not things of unique kind, but things of lots of different kinds.

pedaços. Pode ser frágil sem que seja quebrado. Dizer que é frágil significa que, se alguma vez for, ou tenha sido, golpeado ou forçado, ele se fará, ou se fez, em pedaços. Dizer que o açúcar é solúvel, significa dizer que, se for colocado na água, o açúcar se dissolveria ou se dissolveu (RYLE, 1949, p. 43, tradução nossa)<sup>37</sup>.

As disposições se caracterizam por tornarem possível a aquisição e o desenvolvimento de habilidades de um agente situado e em interação com um contexto. A disposição do vidro de ser frágil está contida em sua própria microestrutura físico-química que, em um ambiente propício, se manifesta. Desse modo, pensar e agir faria parte de um mesmo plano: as ações são as próprias expressões de disposições em constante desenvolvimento.

Nesse sentido, Ryle (1949) critica a tradição intelectualista que estabelece uma dicotomia artificial entre o conhecimento filosófico-científico e o conhecimento comum diretamente relacionado à experiência cotidiana. A crítica de Ryle se fundamenta na medida em que de uma perspectiva intelectualista, o "conhecimento comum" careceria de justificação adequada, supostamente permanecendo no âmbito da opinião. Ryle defende que o conhecimento comum se justifica por si mesmo no momento de sua emergência no contexto. Desse modo podemos dizer que Ryle defende uma postura na qual o conhecimento resultaria das interações agente/mundo, dado o caráter externo e público dos processos mentais, agora considerados indissociáveis das possibilidades de ação entendidas como atualizações de estados disposicionais.

Operar logicamente é um comportamento inteligente adequado a determinadas situações que são vividas por indivíduos. As habilidades de um cirurgião, por exemplo, que permitem *performances* inteligentes não se baseiam na observância de regras metodológicas, mas estão incorporadas no próprio movimento de suas mãos (RYLE, 1949, p. 27). A ação habilidosa pode ser satisfatoriamente descrita enquanto *ação cognitivamente carregada*.

Assim, ações cognitivamente carregadas não são explicadas por meio de causas ocultas, mas sim por meio de *saber como* realizar operações com eficácia, segundo as disposições do agente, adquiridas e aperfeiçoadas pela aprendizagem, efetivadas no plano da ação. Um patinador, ao patinar, embora pense que o gelo sempre resistirá aos seus movimentos, utiliza-se de equipamentos que garantem a sua segurança; isto é, não infere daquilo em que acredita um hábito, pois habituar-se a pensar sempre da mesma maneira (o gelo sempre resistirá) pode colocar sua vida em risco (RYLE, 1949, p. 134). O fato de o

---

<sup>37</sup> When we describe glass as brittle, or sugar as soluble, we are using dispositional concepts, the logic force of which is this. The brittleness of glass does not consist in the fact that it is at a given moment actually being shattered. It may be brittle without ever being shattered. To say that it is brittle is to say that if it ever is, or ever had been, struck or strained, it would fly, or have flown, into fragments. To say that sugar is soluble is to say that it would dissolve, or would have dissolved, if immersed in water.

patinador estar situado no mundo e ter a disposição de considerar outras possibilidades futuras (o gelo poderá quebrar, e isso já aconteceu antes e pode acontecer novamente) faz com que ele tenha um comportamento que supera o hábito mecânico.

Ryle problematiza as dicotomias tradicionais entre mente\corpo, interno\externo, vida privada da mente\vida pública do corpo, sendo estas consideradas como resultado de equívocos lógicos. Para o filósofo, os agentes que possuem disposições para ações habilidosas são incorporados e contextualmente situados. Desse modo, Ryle minimiza a supremacia, e suposta procedência causal, da mente sobre o corpo, tão exaltada pela doutrina internalista. Nesse sentido, Ryle analisa as disposições a partir de um viés duplo: elas não são reduzíveis ao plano físico mas dependem diretamente deste para existir. A fim de endossar o debate acerca da definição de *disposições* apresentaremos abordagens do conceito posteriores à Ryle, como se segue.

Contemporaneamente, surgiu uma nova vertente no estudo das disposições: o denominado *neodisposicionalismo* (MANLEY; WASSERMAN, 2012), cujo viés é a análise das disposições a partir de condicionais lógicas. Nesse contexto, a análise das disposições de um organismo para um determinado comportamento se daria por meio da análise condicional, na forma de sentenças do tipo: “N está disposto a M quando C, se e somente se N faria M se C” (MANLEY; WASSERMAN, 2012, p. 3). Manley e Wasserman propõem a reformulação dessa análise de modo que possa abarcar com maior satisfação a complexidade envolvida no conceito de *disposição*. Para os autores, “N está disposto a M quando C, se e somente se N faria M em uma proporção adequada de casos” (MANLEY; WASSERMAN, 2012, p. 3). Tal tipo de análise é problemática, principalmente porque a possibilidade da existência de um contra-exemplo que falsifica a afirmação de enunciados disposicionais é muito grande. Os próprios autores se mostram descrentes quanto à análise genuína das disposições a partir de condicionais lógicas.

Como exemplo, Manley e Wasserman (2012) apresentam um dispositivo eletrônico associado a fios inicialmente mortos (que não estão conduzindo eletricidade no momento), mas que, ao serem tocados, passam a conduzir corrente elétrica. Assim, os fios conduziram eletricidade somente se tocados, mas não se pode inferir daí que os fios são dispostos a produzirem eletricidade. A interpretação da situação se transfere de: “Se os fios forem tocados conduziram eletricidade”, para: “O fio está disposto a conduzir eletricidade quando tocado”. Para Manley e Wasserman, esse é um contra-exemplo para a simples análise condicional das disposições já que apontam para uma situação que não é o caso necessariamente (“Se os fios forem tocados conduziram eletricidade”), sugerindo que “algo pode ter a disposição de H em

C, mesmo que não seja o caso que seria M em C” (MANLEY; WASSERMAN, 2012, p. 7). Para os autores, isso acontece porque a linguagem comum não articula plenamente as condições de estímulos das disposições que são expressas, mesmo por predicados disposicionais explícitos como: “x está disposto a quebrar-se se vier a cair”. Nesse sentido, o condicional teria que possuir um antecedente que especifique os aspectos relevantes do ambiente em um caso específico, do tipo: “x se quebraria se caísse de uma altura tal, em tal e tal condições”.

A análise das disposições a partir de condicionais aponta para a possibilidade de que disposições possam ser graduais (“x é mais disposto a quebrar-se do que y”, “x é mais disposto a quebrar-se do que a voar”, etc). Tal característica das disposições colabora para a crítica à análise destas a partir de condicionais. Como argumentam, Manley e Wasserman (2012, p. 9-10) uma disposição pode ser gradual, mas uma análise condicional não é gradual independentemente de quão cuidadosamente se especificarem os casos de ocorrência. Isso porque disposições podem ou não se atualizar, na medida em que dependem, além da estrutura físico-química do agente, da estrutura do ambiente.

A questão que colocamos aqui é: modelos mecânicos, uma vez situados contextualmente e atualizando *disposições*, seriam capazes de adquirir e desenvolver *habilidades*, no sentido de Ryle, não previstas por seu programador? Como apontamos anteriormente a aquisição e desenvolvimento de habilidades se dá no plano da ação de acordo com as disposições do agente. Sendo assim, trata-se de um processo não totalmente controlável, pois entre meio e agente são estabelecidas relações dinâmicas, existe um agente que possui determinadas disposições e existe um meio que oferece inúmeras possibilidades de ação. A habilidade de agir do mundo emerge dessa convergência (entre agente e ambiente). Trata-se, então, de um processo emergente cujo produto final não pode ser totalmente determinado, pois não há um controlador central que dita o percurso da relação estabelecida entre o agente e o ambiente. Nesse sentido, podemos caracterizar o processo de aquisição e desenvolvimento de habilidades como auto-organizado. Assim, os modelos computacionais teriam que estar situados em um meio real (em oposição ao puramente virtual ou fisicamente controlado) e, além disso, possuir disposições efetivas para agir *habilmente* no referido meio, estabelecendo relações auto-organizadas, ou seja, sem um controlador central que dita o desenvolvimento do processo. Exporemos na sessão seguinte os pressupostos da Teoria da Auto-organização, segundo Debrun (2009), a fim de apresentar mais um aspecto significativo das habilidades e, posteriormente, como este processo se daria na modelagem computacional. A auto-organização, nesse sentido, estaria no cerne da aquisição e desenvolvimento das

habilidades e por tanto seria um aspecto basilar da modelagem computacional que visa projetos de modelos habilidosos.

#### 2.4 A TEORIA DA AUTO-ORGANIZAÇÃO E AS HABILIDADES DISPOSICIONAIS

Apontamos na sessão anterior que o processo de aquisição de habilidades a partir de disposições de um agente pode ser caracterizado como um processo auto-organizado, pois nesse processo não existe um controlado único e absoluto que o determine. Trata-se de uma constante troca entre agente e meio no plano de ação, de acordo com as disposições que ambos possuem e oferecem.

Nesse sentido, segundo Debrun (2009, p. 54) “Há auto-organização cada vez que o advento ou a reestruturação de uma forma, ao longo de um processo, se deve principalmente ao próprio processo – a características nele intrínsecas-, e só em grau menor às suas condições de partida, ao intercambio com o ambiente ou à presença eventual de uma instância supervisora”. Desse modo, quanto maior o hiato entre o estado inicial e a forma final do processo, maior é o grau de auto-organização. E como ainda aponta Debrun (2009, p. 55) “A auto-organização é sempre em algum grau uma criação”, no sentido de que sempre surgem características que não estavam presentes inicialmente no processo, mas que acabam constituindo a forma final deste. É uma criação, mas continua sendo um processo que se desenvolve ao longo do tempo. E justamente por ser um processo não é mera decorrência das condições iniciais. Nas palavras de Debrun:

[...] o processo de auto-organização apenas “herda” esse começo, que ele vai levar em conta de modo muito variável. O começo é importante porque introduz um corte com o passado e com o contexto – o que permite ao processo em conjunto tornar-se independente, em parte, do “resto do universo” (DEBRUN, 2009, p. 56)

O corte com o passado é importante pois é através dele que os agentes que instanciam processos auto-organizados se tornam elementos *realmente distintos* e assim são livres, no sentido de que ainda não têm suas possibilidades de ação limitadas pela organização que poderá ser posteriormente criada, estando apenas submetidos às leis gerais da natureza. Como aponta Debrun (2009, p. 57) “Isso [a ausência de conexões] faz com que esses elementos, em vez de se *condicionarem* (um ao outro, ou um e outro reciprocamente), se ‘encontrem’, ficando livres, portanto, para conexões novas, inauditas – que vão surgir ‘aqui e agora’ – e

não apenas para se atualizar ou revelar.”. É nesse contexto, de elementos livres e soltos, que a novidade pode ocorrer, acarretando em uma nova propriedade, configuração, comportamento, ação do sistema, etc. Esse processo ocorre por si mesmo, em um trabalho de si sobre si, de maneira autônoma às condições iniciais que é favorecido pelas próprias condições iniciais ou por parte delas (DEBRUN, 2009, p. 58).

As condições iniciais, embora não determinantes, seriam propiciadoras de uma certa autonomia. Nesse sentido, em processos auto-organizados a natureza dos elementos envolvidos é relevante, mas não determinante: deve haver um corte com o passado, tornando os elementos livres para se organizarem mas, simultaneamente, eles devem estar sob condições de partida favoráveis para que o processo seja auto-organizado. No caso do desenvolvimento de habilidades, por exemplo, o processo em si é auto-organizado pois diz respeito à um trabalho de partes autônomas a partir de si mesmo, mas as condições de partida (configuração do sistema, condições do meio, etc) desempenham um papel importante no produto final. Nada, nem ninguém determinaram no nascimento de cada ser humano que ele efetivaria a capacidade de calcular (visto que algumas pessoas enfrentam sérias dificuldades para calcular). As condições iniciais (estrutura biológica, constituição física e contexto) propiciaram o desenvolvimento desta capacidade. Não é que a capacidade de calcular é auto-organizada, mas o processo em que ela se dá é *auto*, é *sobre si mesmo*, não é previamente determinada (embora muitos fatores favoreçam, ou não, seu desenvolvimento), tanto que existem seres humanos incapazes de calcular. A *forma*, nesse sentido, não é a resultante passiva do processo. Como não existe um controlador central, a forma (por exemplo, um sistema com a capacidade de calcular) se dá por meio de ajustes organizacionais e como produto de um processo legitimamente auto e por si só se configura como uma *identidade*. Sob esta abordagem, o ser humano que possui a capacidade de calcular, a possui enquanto uma capacidade emergente de partes realmente distintas e livres que se organizaram no espaço e no tempo de modo a conferir a este indivíduo tal capacidade. Esse ser humano é, portanto, *calculador* no sentido de que esta capacidade passa a fazer parte de sua identidade pessoal. Na auto-organização, portanto, os elementos se encontram, pois como aponta Debrun (2009, p. 119): “[...] elementos que não se condicionam entre si só podem se encontrar [...]”. Ao contrário do que ocorre, por exemplo, na área lógico-matemática, onde dois teoremas não podem se encontrar: são conectados de antemão entre si por axiomas, regras de derivação etc.”.

A auto-organização, nesse sentido, possui como motor principal a própria interação entre elementos realmente distintos ou entre partes semi-distintas que constituem um sistema

(DEBRUN, 2009, p. 59). Assim, tal como caracterizada por Debrun, a auto-organização pode ser de dois tipos: auto-organização primária e auto-organização secundária. A auto-organização é primária quando: “[...] ela não parte de uma ‘forma’ (ser, sistema etc.) já constituída, mas que, ao contrário, há ‘sedimentação’ de uma forma.” (DEBRUN, 2009, p. 60). E a auto-organização é considerada do tipo secundária na medida em que não parte de simples elementos, mas de um ser ou sistema já constituído, que “decide” e controla o processo rumo a uma complexidade maior. Nas palavras do filósofo:

No concernente ao que ocorre no organismo diremos que há auto-organização secundária quanto esse organismo consegue passar, a partir de suas próprias operações, exercidas sobre ele próprio, de determinado nível de complexidade – corporal, intelectual, existencial – para um nível superior. A auto-organização aqui é secundária à medida que ela não parte de simples elementos, mas de um ser ou sistema já constituído (DEBRUN, 2009, p. 61).

A auto-organização primária não desempenha uma função, mas trata-se de um desenvolvimento “selvagem”. Quando o processo se encontra mais ou menos consolidado ele comporta “autofunções”, ou seja, funções exercidas em benefício do próprio processo. Mesmo no caso da auto-organização secundária, no qual um sistema já está constituído, as funções desempenhadas não são definidas ou implantadas por outrem: em ambos casos de auto-organização, o processo é sempre um trabalho de si sobre si. Dessa caracterização de auto-organização pode-se perguntar pelo sujeito. A auto-organização é um processo sem sujeito? E na auto-organização secundária, quem é esse sistema já constituído que instancia o processo? Como afirmar que o desenvolvimento de habilidades é um processo auto-organizado e que parte de disposições de um agente? Como apontado por Debrun, parte dos estudos acerca da auto-organização tem se desdobrado para reflexões além de um sujeito, pois “O importante, para eles [os estudiosos da Teoria da Auto-organização], é apreender a natureza, e alguns artefatos, como sendo ‘auto’, ou capazes de se tornarem ‘auto’, mas não forçosamente ‘sujeito’.” (DEBRUN, 2009, p. 69). Essa afirmação de Debrun é importante, pois ela desloca o eixo das discussões acerca da auto-organização da estrutura de um sujeito. Ou seja, a auto-organização se refere à maneira como um processo se dá e tal processo não é necessariamente limitado a um sistema de um tipo específico.

A noção de sujeito na auto-organização pode permanecer, mas não existe um sujeito *específico*, *onipotente* e *absoluto* nesse processo que atue como causa determinante. Duas situações podem ocorrer, como aponta Debrun (2009, p. 69): “Ou o sujeito está ausente da auto-organização, havendo quando muito – a nível físico, químico ou biológico – uma

subjetividade ‘difusa’ [...]. Ou então o sujeito está presente, mas apenas como elemento [...] ”. O sujeito seria apenas mais uma parte de um processo muito maior. O relevante aqui é que o processo de auto-organização ocorre em um espaço com uma temporalidade e seu desenvolvimento é definido no momento presente. Tais condições parecem, segundo Debrun, ser independentes da determinação condicionante de um sujeito. O processo de auto-organização diz respeito às condições de partida e não à natureza dos elementos que o propiciam. Por exemplo, para que a capacidade de calcular emergisse (de forma auto-organizada, pois não existe um centro controlador do processo), as condições de partida necessitam ser favoráveis, independente de um sujeito (independente se o sujeito for um humano ou uma máquina).

Tendo em vista a natureza mecânica dos modelos, parece que a própria estrutura mecânica e formal teria que ser constituída de tal maneira que, a partir desse substrato mecânico, pudessem emergir, por aprendizagem, habilidades disposicionais a partir do meio. Desse modo, o modelo mecânico seria capaz de adquirir e aperfeiçoar suas habilidades através da aprendizagem e de sua interação com o ambiente. Segundo os estudiosos da Teoria da Auto-Organização, o processo seria possível independentemente das condições materiais envolvidas, pois o cerne do processo é o próprio desenvolvimento. A auto-organização implica haver trocas (informacionais, energéticas, materiais) entre elementos livres e o meio em que estão inseridos. No que diz respeito à modelagem computacional, Brooks aponta que o objetivo da nova IA deveria ser o “(...) de construir agentes móveis completamente autônomos que coexistissem no mundo com os seres humanos e fossem vistos por estes seres humanos como entidades inteligentes por si mesmas” (BROOKS, 1991a, p. 142). Abordaremos o projeto de da nova IA proposto por Brooks no próximo Capítulo.

**CAPÍTULO 3 - A NOVA ROBÓTICA: UMA REFLEXÃO ACERCA DOS  
PARADIGMAS CONTEMPORÂNEOS DA IA**

*We alone on this Earth have syntax, and we alone on this Earth have technology  
(BROOKS, 2002, p. 05)*

## APRESENTAÇÃO

No Capítulo 1 desta dissertação nos dedicamos a expor os pressupostos basilares que norteiam o ramo da Ciência Cognitiva denominada de Inteligência Artificial. No Capítulo 2, a partir do histórico traçado anteriormente, apontamos para teorias explicativas acerca dos processos mentais que se baseiam em um paradigma situado e incorporado. Dentre os processos mentais, nos apegamos ao conceito de *inteligência*. Até o presente momento nos pautamos nas seguintes questões: O que é ser inteligente? Quais critérios devem ser satisfeitos para que um sistema possa ser considerado inteligente? Seria a inteligência uma propriedade ou qualidade exclusivamente biológica? Tendo essas interrogações em vista, no Capítulo 3 apresentamos uma abordagem de modelagem computacional que considera comportamentos inteligentes aqueles que são executados no plano da ação e envolvem interações bem sucedidas com o mundo. Sendo assim, tal abordagem rompe com o paradigma tradicional que apresentamos no Capítulo 1 e se aproxima de pressupostos da Teoria da Cognição Situada e Incorporada e dos princípios da Teoria da Auto-Organização que apresentamos no Capítulo 2. Para dar continuidade à discussão, trataremos, na **seção 3.1**, do projeto de Nova Robótica de Rodney Brooks segundo o qual os comportamentos são considerados de maneira *incremental* e o contexto real não controlado tem um papel determinante nesse processo. Na **seção 3.2**, a partir dos modelos construídos por Brooks, apontaremos que o processo de incrementalização pode relacionar-se com a auto-organização. Nesse sentido, apresentaremos pesquisas que fundamentam esta perspectiva como as de Beer (1996, 2003). Na **seção 3.3**, com base na ideia de que processos cognitivos se desenvolvem de maneira incremental e, possivelmente, auto-organizada, apontaremos que os agentes envolvidos nesse processo podem ser considerados sistemas complexos participantes de um processo carregado cognitivamente que envolve aprendizagem e que é desenvolvido no âmbito das ações. Por fim, na **seção 3.4** analisaremos como a informação atua no processo de complexificação de comportamentos e que papel ela desempenha em modelagens que visam desenvolver agente situados, incorporados, auto organizados e habilidosos.

### 3.1 O PROJETO DE BROOKS PARA UMA NOVA ROBÓTICA

Como apontado por Brooks (1991b), o objetivo da IA tradicional consiste em “tentar construir computadores que façam algo que, quando feito por seres humanos, é indicador de inteligência<sup>38</sup>”. Todavia, tendo em vista os problemas dos modelos criados pela IA tradicional, Brooks (1990, 1991a, 1991b) crítica os pressupostos basilares desse programa de pesquisa. Nesse contexto, a crítica de Brooks se detém primordialmente no papel reducionista que a representação desempenha na modelagem computacional. Segundo Brooks (1998), Newell & Simon instauraram a hipótese de que o ser humano utiliza um sistema puramente físico e simbólico para pensar (tal como apontamos no Capítulo 1 deste trabalho). Tal hipótese se tornou implícita e corrente no ramo da IA e dela decorre que o sistema depende de representações internas, explícitas e uniformes para agir e que sua ação depende de um modelo prévio de mundo. Esse paradigma é dominado por problemas tanto de acesso a fatos relevantes que serviriam de parâmetro de ação do modelo quanto de aplicabilidade. A dificuldade que envolve o paradigma de modelagem computacional que se baseia exclusivamente em representações consiste em determinar o critério de seleção de características do mundo relevantes para serem representadas e, uma vez escolhidas as características relevantes, o problema é como representa-las, como instanciá-las adequadamente no modelo<sup>39</sup>.

Para o roboticista “A ideia [da IA tradicional] era de que, ao representarmos explicitamente apenas os fatos relevantes faríamos com que a semântica de um mundo (muito complexo em sua superfície) fosse reduzida mais uma vez a um sistema fechado” (BROOKS, 1991a, p. 141). A representação, ao ser instanciada em modelos robóticos a partir de ferramentas computacionais, reduz o conceito de *inteligência* à manipulação abstrata de certas informações consideradas relevantes. Como ainda apontado por Brooks (1991a), a manipulação de informações nos modelos da IA tradicional se dá em um sistema central com módulos perceptuais (que “captam” entradas – *inputs*) e módulos de ação (*outputs*). Nesse

<sup>38</sup> [...] tries to make computers do things that, when done by people, are described as having indicated intelligence.

<sup>39</sup> Brooks aponta que a IA neo-clássica adiciona ideias bayesianas e outras ideias probabilísticas ao paradigma tradicional de modelagem da IA. Ao contrário da estatística tradicional que trabalha com um espaço amostral fixo, a probabilidade bayesiana comporta a inserção de novas informações no cálculo inferencial. Trata-se de uma probabilidade condicional que conforme a inserção de novas informações diminui o espaço amostral, pois está atrelado ao acontecimento de outro fato que já ocorreu. Por exemplo, segundo a estatística tradicional a probabilidade de se tirar um ás do meio de um baralho completo é de 4/52 (o baralho completo é constituído por 52 cartas e existem 4 ases, 1 de ouro, 1 de paus, 1 de espada e 1 de copas). Se condicionarmos o espaço amostral à: “ás de ouro” a probabilidade seria de 1/13 (52 cartas no total, 4 naipes possíveis:  $52/4 = 13$ ). Existe apenas um ás de cada naipe, logo 1/13). Nesse contexto, a inserção de uma nova informação (no sentido de que tal informação não estava presente nas condições iniciais), altera o comportamento do sistema que passa a trabalhar com um novo espaço amostral que determinará seus passos seguintes.

contexto, “Os módulos perceptuais apresentam uma descrição simbólica do mundo e os módulos de ação consideram uma descrição simbólica de ações desejadas, fazendo com que elas aconteçam no mundo” (BROOKS, 1991a, p. 153). A representação seria o elemento mediador entre o que o módulo perceptivo apreende e a ação desencadeada por ele. Assim, o modelo computacional desempenharia *funções* de acordo com o código de programação escrito por seu programador. No entanto, segundo Brooks (1991a), o paradigma funcional adotado pela IA tradicional é frágil, na medida em que é necessária uma grande cadeia de módulos funcionais para que o modelo seja capaz de simular algum comportamento. A grande quantidade de módulos requeridos facilita o surgimento de uma falha operacional fatal.

Em contrapartida à abordagem da IA tradicional, Brooks (1991a) sugere um paradigma que se pautem em *atividades*. Para Brooks (1991a, p. 153) “uma atividade é um padrão de interações com o mundo”, ou seja, o modelo computacional não se reduziria apenas a manipulações simbólicas regidas por representações. Assim, o modelo, ao desempenhar uma atividade, interage diretamente com o meio em que está inserido. Desse modo, Brooks (1998) argumenta que a inteligência é resultado direto de cinco características interligadas: desenvolvimento organizacional (*developmental organization*), interação social (*social interaction*), incorporação (*embodiment*), acoplamento físico (*physical coupling*) e interação multimodal (*multimodal integration*). O desenvolvimento organizacional diz respeito à capacidade de adquirir com sucesso habilidades e competências complexas. A interação social permite aos seres humanos, por exemplo, explorar outros seres para aprendizado e conhecimento. A incorporação e o acoplamento físico possibilitam aos humanos utilizarem o próprio corpo e o mundo como ferramentas para a organização e manipulação de conhecimento. Por fim, a integração multimodal (integração dos vários módulos dos quais os humanos são compostos: tato, fala, visão e olfato, por exemplo) faz com que a precisão e a eficácia dos sistemas sensorio e motor sejam maximizadas. Desse modo, a partir do funcionamento do comportamento humano, Brooks (1998) argumenta que sistemas construídos baseados no desenvolvimento, na interação, na incorporação e acoplamento físico e na interação multimodal possuem a capacidade de *simplificar* comportamentos considerados complexos. Assim, a aprendizagem seria facilitada, pois constituiria um processo constante de criação de mecanismos que podem ser utilizados em tarefas complexas que são requeridas pelo ambiente.

A metodologia de Brooks se assenta basicamente no princípio de que comportamentos complexos podem derivar de comportamentos mais simples. Tal postura dá margem à argumentação de que os humanos tendem a minimizar a sua representação interna de mundo e

focalizam apenas nas que são *imediatamente* relevantes. O comportamento do sistema, nesse contexto, não se baseia em um modelo prévio de mundo, mas o *cria* no momento presente a partir das condições consideradas relevantes. Trata-se de desenvolver um esquema de mundo atual condicente com o objetivo dado ao modelo computacional. Desse modo o modelo poderia agir no mundo com uma direcionalidade, lidar com mudanças em seu ambiente dinâmico, resistir ao seu ambiente, manter múltiplos objetivos e mudá-los adaptando-se. A proposta de Brooks consiste em construir esquemas de mundo que serviriam de guia para os comportamentos e ações dos modelos computacionais, não assentados em representações prévias do mundo.

Todavia, parece que nessa abordagem a representação prévia de aspectos do mundo não desaparece totalmente, mas, na melhor das hipóteses, tem apenas seu papel descentralizado. Em outras palavras, os modelos construídos por Brooks possuem um tipo de representação de certos aspectos do mundo (instanciadas por seu programador e limitadas também por sua arquitetura física), mas ela não desempenha o papel central e determinante no desencadeamento dos comportamentos. A representação, nesse sentido, se configuraria como uma das condições iniciais que propiciariam certo tipo de comportamento do modelo. Como apontado por Brooks:

Trabalhos recentes dentro desta [nova] abordagem têm mostrado o uso de representações, expectativas, planos, metas e aprendizagem, mas sem recorrer aos usos tradicionais de representações centrais abstratamente e simbolicamente manipuladas. A percepção dentro destes sistemas é muitas vezes um processo ativo e a dinâmica das interações com o mundo é extremamente importante (BROOKS, 1991b, p. 1227)<sup>40</sup>

Para levar a cabo seu projeto de modelagem computacional, Brooks também sugere uma metodologia de engenharia específica para tornar possível a construção de modelos computacionais que interagissem diretamente com o meio. A metodologia de Brooks consiste na construção de modelos computacionais a partir de uma *arquitetura de subsunção*. Segundo esta arquitetura, os modelos seriam construídos em camadas reativas, ou seja, cada camada física constituinte do modelo estaria diretamente relacionada com o meio e decidiria quando agir. Para Brooks (1991a, p. 153) esse tipo de camadas, “... não se trata de sub-rotinas que estão dependentes de alguma outra camada e que são evocadas por ela”. As camadas são reativas por si mesmas, dada a maneira específica como foram construídas e também a

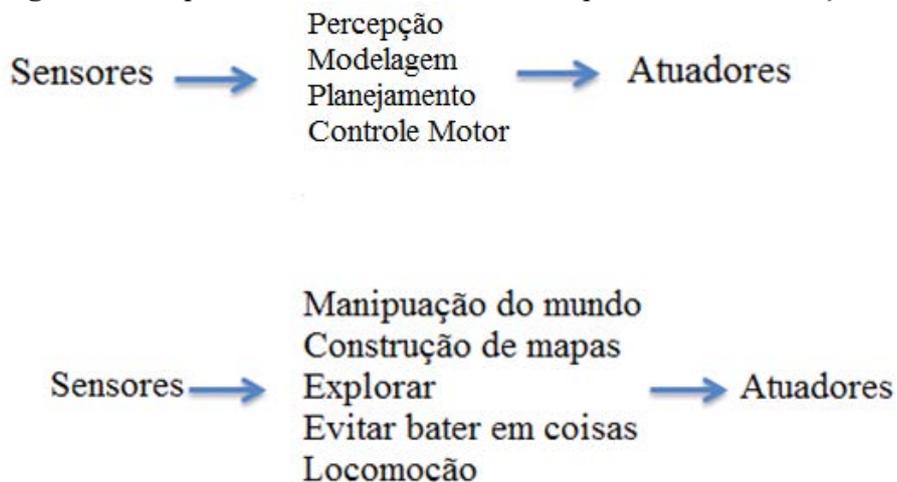
---

<sup>40</sup> Recent work within this approach has demonstrated the use of representations, expectations, plans, goals, and learning, but without resorting to the traditional uses, of central, abstractly manipulable or symbolic representations. Perception within these systems is often an active process, and the dynamics of the interactions with the world are extremely important.

amplitude e flexibilidade da linguagem de programação utilizada no desenvolvimento do código-fonte do modelo.

A arquitetura de subsunção proposta por Brooks se baseia na construção de camadas físicas constituídas de elementos computacionais simples, mas amplos o suficiente para que possam reagir diretamente aos estímulos do ambiente. Por exemplo, um dado modelo é capaz de se mover em um ambiente real, desviando de possíveis obstáculos. Seguindo o paradigma da arquitetura de subsunção, o modelo seria constituído de camadas físicas reativas responsáveis por comportamentos *simples*. A camada 1 poderia ser capaz de mover-se para frente, a camada 2, de mover-se para a direita, a camada 3 por elevar uma perna e a 4 de parar os movimentos. Cada camada reage diretamente à configuração do meio percebida pelo modelo. Nesse contexto, ao perceber um obstáculo em seu caminho, o modelo imediatamente suprime a camada 1 (mover-se para frente) e ativa a camada 2, por exemplo. A figura 4 apresenta a estrutura geral da arquitetura de subsunção.

**Figura 4** - Esquema do funcionamento da arquitetura de subsunção.



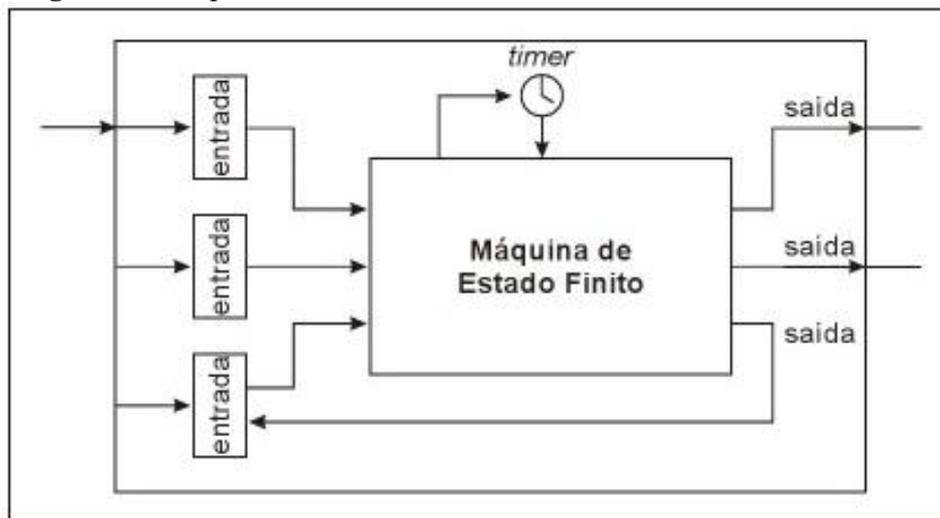
**Fonte:** Brooks (1991b, p. 1229, tradução nossa).

Note-se que o sensor de informação (aquele que recebe os dados do ambiente em que o modelo está situado) está diretamente relacionado com o módulo atuador. Não existe, segundo Brooks, uma mediação entre percepção e ação nos modelos construídos por meio da subsunção. Como apontado pelo roboticista: “Ao invés de modularizar a percepção, o modelo de mundo, o planejamento e a execução, a nova abordagem constrói sistemas de controle inteligente, em que cada módulo individual gera diretamente alguma parte do comportamento do robô” (BROOKS, 1991b, p. 1227). Uma das dificuldades da modelagem da IA tradicional diz respeito à decomposição de comportamentos e habilidades em funções. Assim, como apontamos, na IA tradicional, para que o modelo seja capaz de realizar um comportamento

sofisticado, seria necessária uma extensa cadeia de módulos funcionais, os quais, devido à intensa manipulação de informações, dão margem a erros fatais de processamento. Isso impede que o modelo seja capaz de responder a um mundo dinâmico, não controlado. A proposta de Brooks, por outro lado, tem como objetivo desenvolver modelos computacionais com ligações curtas entre os sensores perceptuais (aqueles que percebem a informação presente no ambiente) e atuadores (módulos que desencadeiam o comportamento *final* do modelo), tornando possível, pelo menos em princípio, responder rapidamente às mudanças do mundo.

Para que a ligação entre os sensores do modelo seja curta e eficaz para responder ao mundo, segundo a abordagem de Brooks, cada camada é uma máquina de estado finito ampliada (MEFA). O funcionamento desta máquina se baseia na entrada de dados (*input*) e a saída de outros dados (*output*) a partir de uma função de transição. A figura 5 ilustra o funcionamento de uma MEFA.

**Figura 5** – Esquema do funcionamento de uma MEFA



**Fonte:** Teixeira (1998. p. 136).

Nesse contexto, como mostra a figura 5, um *output* pode se tornar um *input* de outra máquina ou de si mesma. Este *output* por sua vez pode acionar ou inibir a realização da função determinada. Assim, um conjunto de máquinas de estado finito ampliada constitui uma função comportamental como *locomover-se*, *erguer a perna*, *pegar objetos*. O modelo assim construído, e tendo como referência o próprio ambiente em que está inserido, irá *decidir* quando ativar ou inibir determinada camada que carrega em si uma determinada função comportamental. O modelo seria autônomo, nesse sentido específico, para tomar uma decisão.

A ideia chave da nova robótica de Brooks é a construção de agentes autônomos, no sentido apontado, configurados fisicamente como um robô móvel que realiza algumas tarefas úteis em um ambiente que não foi especialmente estruturado e adaptado para ele. O modelo computacional de Brooks é instanciado fisicamente na forma de um robô, pois, como apontamos anteriormente, a corporeidade é importante para o processo de incrementalização de comportamentos. É por meio de um *corpo* que o modelo pode se situar em um ambiente, incorporar dele informações relevantes e se posicionar no espaço-tempo. No caso dos modelos computacionais, o corpo possível é físico e estruturado segundo a arquitetura de subsunção apresentada. Desse modo, “Os robôs estão situados em um mundo no qual eles não lidam com descrições abstratas, mas com o ‘aqui’ e ‘agora’ do meio ambiente que influencia diretamente o comportamento do sistema” (BROOKS, 1991b, p. 1227).

Segundo Brooks (1991b, p. 1229), os comportamentos do modelo computacional construído a partir da arquitetura de subsunção são sofisticados na medida em que novas camadas reativas são adicionadas. Como as camadas são independentes entre si, a possibilidade de um erro fatal referente ao processamento de informações relevantes é minimizado, pois cada camada lida apenas com um conjunto determinado e reduzido de dados. Nesse sentido:

Cada uma das camadas é uma peça produtora de comportamento em sua própria rede, embora possa implicitamente contar com a presença de partes de redes anteriores. Por exemplo, a camada cuja função é explorar, não precisa explicitamente evitar obstáculos, pois a arquitetura sabe que a camada existente para essa função vai cuidar disso<sup>41</sup>. (BROOKS, 1991b, p. 1227)

Além do mais, um esquema de prioridades é desenvolvido para que os possíveis conflitos de comportamentos sejam resolvidos. É instanciado no modelo, por exemplo, se, ao perceber um obstáculo, ele deve primeiro parar seu movimento contínuo para frente ou se deve primeiro erguer uma perna. Nesse contexto, a autonomia do modelo defendida por Brooks se resume à possibilidade (previamente limitada) de acionar uma ou outra camada reativa da arquitetura. A arquitetura de subsunção destoa das utilizadas pela IA tradicional no sentido de que não existe um modelo central de mundo explicitamente representado no interior dos robôs. Tampouco existe uma separação implícita de dados e computação, pois ambos foram distribuídos dentro das redes de ação em cada camada. A fronteira entre a

---

<sup>41</sup> Each of the layers is a behavior-producing piece of network in its own right, although it may implicitly rely on the presence of earlier pieces of network. For instance, an explore layer does not need to explicitly avoid obstacles, as the designer knows that the existing avoid layer will take care of it.

computação e o mundo, segundo Brooks, é mais complicada de ser delimitada, pois os robôs construídos segundo a arquitetura de subsunção dependem fortemente da dinâmica das interações com o mundo para o desenvolvimento dos comportamentos.

Brooks, enquanto esteve no Laboratório de Inteligência Artificial do MIT, construiu alguns modelos computacionais a partir dos pressupostos da arquitetura de subsunção. Os modelos computacionais de Brooks, na verdade, são denominados de robôs, pois são instanciados fisicamente em *corpos* para que o contato com o meio seja efetivo. Nesse sentido, os robôs de Brooks estão para além do mero funcionamento em um *software*, eles estão *situados e incorporados* em um meio.

No início dos anos 80, Brooks desenvolveu o robô Allen. O robô foi construído a partir da arquitetura de subsunção, cujas camadas eram majoritariamente reativas e cujos módulos perceptuais permitiam a interação com o meio. Allen foi construído com 3 camadas. A primeira camada possibilita ao robô evitar obstáculos dinâmicos e estáticos presentes no meio. A segunda camada faz com que o robô a cada 10 segundos tenha o *desejo* de se movimentar de modo a explorar o mundo em que está situado. A terceira camada possibilita ao robô *perceber* o ambiente amplamente de modo a detectar as potencialidades de ação nele contidas. Com a subsunção das três camadas, Allen, portanto, era capaz de perceber o ambiente de modo a ter informações que permitiam o planejamento de suas ações. Percebendo o mundo em que está situado, Allen era capaz de evitar obstáculos que impediam a realização de sua meta (mover-se e explorar o mundo).

Posteriormente, Brooks desenvolveu o robô Hebert que, além de perceber o mundo em que estava situado, movendo-se evitando obstáculos, era capaz de pegar latas de refrigerantes em cima de mesas. Já o robô Genghis foi desenvolvido para mover-se em ambientes de terrenos acidentados (BROOKS, 1990). Nota-se que os robôs de Brooks, até então, foram desenvolvidos a partir da instauração de uma meta geral (mover-se, explorando o mundo em que estão situados) que possibilita ao modelo *experienciar* o ambiente. A partir desse paradigma, Brooks desenvolve Cog, um robô humanoide construído para interagir em um ambiente real, não controlado e com a presença de humanos. Como aponta Brooks, Cog foi construído com o objetivo de “explorar questões de desenvolvimento estrutural, incorporação física, integração de múltiplos sistemas sensoriais e motores e interação social <sup>42</sup>” (BROOKS, 1998, p. 01). Com o robô Cog, Brooks pretende levar a cabo a instanciação da hipótese segundo a qual os comportamentos podem se sofisticar de maneira incremental, ou seja, comportamentos considerados simples que se tornam mais complexos a partir da interação

---

<sup>42</sup> To explore issues of developmental structure, physical embodiment, integration of multiple sensory and motor systems, and social interaction

com o ambiente. Para isso, Cog possui um conjunto de sensores e atuadores que simulam a dinâmica sensorio-motora humana, além de um sistema de reconhecimento de faces que lhe permite interagir com pessoas.

Para interagir socialmente, Cog deve compreender as convenções sociais que as pessoas usam e entendem diariamente (Brooks et al, 2000, p. 1). Em outras palavras, o robô Cog deve compreender a lógica do contexto social humano, extrair as regras de convivência e manipulá-las de modo satisfatório e adequado. Nesse contexto, o objetivo de Brooks consiste em desenvolver Cog de tal modo que ele seja capaz de se situar em ambientes naturais autonomamente. A tentativa de modelar a autonomia<sup>43</sup> implica em uma estruturação específica do *designer* e a incorporação física dos sistemas de controle dos robôs, ou seja, o modelo precisa ser desenvolvido de modo a propiciar a disposição de interagir, também do ponto de vista físico. A partir desse pressuposto, os modelos desenvolvidos por Brooks possuem sensores que captam informações visuais do ambiente e as levam até as centrais de processamento que causam as respostas motoras. Tais sensores estão acoplados em uma estrutura física que simula um *corpo*. Além dos sensores, Cog possui rodas para se locomover, braços e uma “cabeça” dotada de “rosto”. Toda essa estrutura é desenvolvida a fim de possibilitar a interação autônoma de Cog com o ambiente e com as pessoas nele inseridas.

Nesse sentido, a proposta de Brooks consiste em dotar Cog das disposições físicas e lógicas para ser autônomo, situá-lo em um ambiente natural (não controlado) e deixá-lo que ele próprio incremente seus comportamentos e esquemas de ação a partir de suas próprias disposições e relações estabelecidas com o meio. A figura abaixo, por exemplo, mostra as expressões faciais de Cog. Segundo Brooks (1998), Cog “acessaria” cada umas das expressões de acordo com o contexto e elas próprias seriam desenvolvidas a partir dele. Por exemplo, se Cog participasse de uma festa ele acessaria a expressão *happiness* e suas estruturas físicas se configurariam de acordo com esta expressão. Ao conviver com pessoas, Cog entenderia que normalmente se fica feliz em festas e que quando as pessoas ficam felizes seus olhos ficam abertos e sua boca arqueada, como em um sorriso. Cog simularia esta configuração a partir da regra social que aprendeu. O aprendizado, nesse sentido, se daria na intersecção entre a configuração previamente estabelecida no robô (sua arquitetura mecânica e

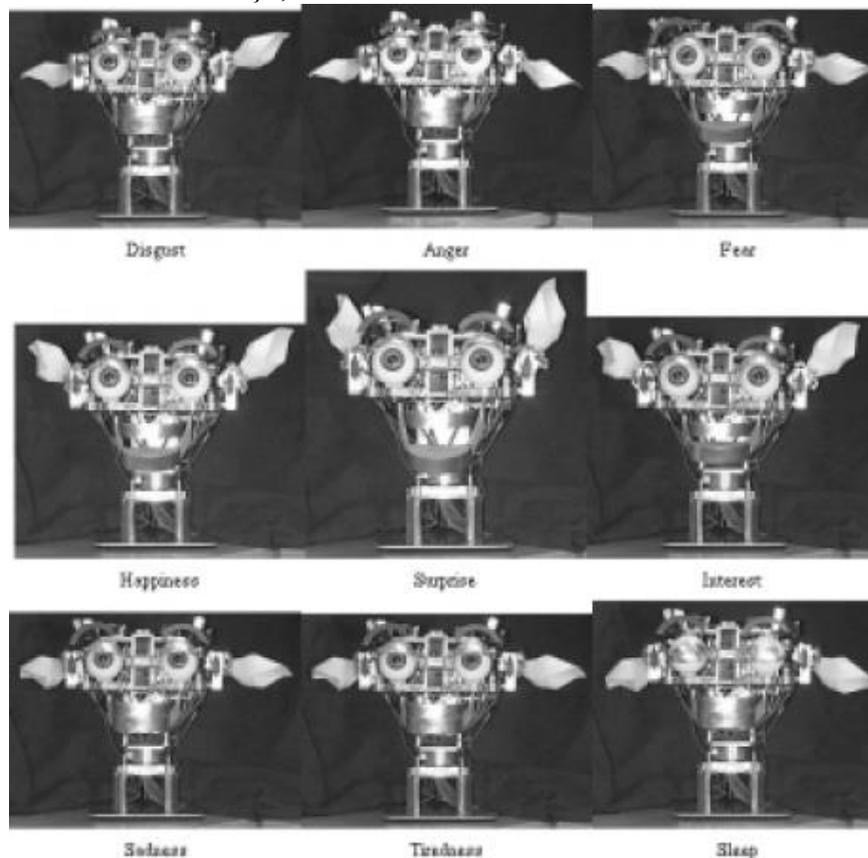
---

<sup>43</sup> A autonomia, tal como caracterizada neste trabalho, seria também um tipo de capacidade inteligente pois envolve a capacidade de tomar decisões a partir de experiências passadas e habilidades adquiridas. Tal capacidade de natureza disposicional necessita da troca de informação significativa com o meio.

lógica de programação) e a experiência vivida no ambiente não controlado no qual está situado.

Assim, os modelos de Brooks se configuram como uma tentativa de entender características humanas naturais (como a autonomia). Não se trata, então, de apenas simular capacidades mas sim de compreendê-las: Cog é também uma *experiência de pensamento*, no sentido de que “Nós esperamos não apenas criar robôs inspirados em capacidades biológicas, mas também contribuir para dar forma e refinar nosso entendimento dessas capacidades.” (Brooks et al, 2000, p.01).

**Figura 5** – Expressões faciais do robô Cog: asco, raiva, sentir (quando toca algo), felicidade, surpresa, interesse cansaço, dormindo.



**Fonte:** Brooks *et al.*. (1998, p. 64).

É importante ressaltar aqui que o projeto de Brooks pretende simular a autonomia, mas parece não focalizar as estruturas nas quais ela se instância, o que nos remete ao projeto funcionalista dos primórdios da IA. O que Brooks deseja é que os modelos sejam autônomos, tal como as pessoas, independentemente do material que os constitui. No caso, os modelos são desenvolvidos a partir de circuitos elétricos, placas de metal, etc., pois é o material disponível para tal no atual estágio tecnológico.

Se analisada superficialmente, a proposta de arquitetura de subsunção, com sua rede de máquinas simples, pode ser confundida com abordagens mecanicistas, tais como o conexionismo e as redes neurais. Brooks (1991a, p. 147) faz questão de apontar que seu projeto de modelagem não é compatível com os pressupostos mecanicistas. Segundo Brooks:

Os conexionistas tentam fazer redes de processadores simples. Assim, as coisas que eles constroem (apenas em simulação, pois nenhum conexionista jamais colocou nenhum robô de verdade em um ambiente real, não importa quão simples) são similares às redes de subsunção que construímos. Entretanto, seus nós de processamento tendem a ser uniformes, e o que eles buscam (conforme seu nome) são revelações a respeito de como podem ser conectados corretamente (algo que, geralmente significa realizar conexões ricas)<sup>44</sup> (BROOKS, 1991a, p. 147)

A proposta de Brooks destoa do conexionismo na medida em que os nós na arquitetura de subsunção são máquinas de estado finito ampliadas diferentes umas das outras e a densidade das conexões é muito menor e não uniforme. Brooks ainda afirma que os conexionistas estão preocupados com a emergência espontânea de representações explícitas nas redes. Para o roboticista: “Nós não temos esperanças desse tipo, porque acreditamos que as representações não são necessárias, e aparecem apenas no olho ou na mente do observador<sup>45</sup>” (BROOKS, 1991a, p. 147). Nesse viés, Brooks distancia seu projeto da abordagem das redes neurais, pois:

A abordagem por redes neurais é a disciplina original da qual o conexionismo é a encarnação mais recente. As pessoas que trabalham com redes neurais alegam que há alguma significação biológica em seus nós de redes: é como se eles fossem modelos de neurônios<sup>46</sup> (BROOKS, 1991a, p. 147).

Segundo Brooks, a maioria dos modelos computacionais baseados em redes neurais parece implausível, pois o número de conexões presentes nesses modelos é muito reduzido se comparado ao número de conexões que realmente existem entre os neurônios. Além do mais,

---

<sup>44</sup> Connectionists try to make networks of simple processors. In that regard, the things they build (in simulation only—no connectionist has ever driven a real robot in a real environment, no matter how simple) are similar to the subsumption networks we build. However, their processing nodes tend to be uniform and they are looking (as their name suggests) for revelations from understanding how to connect them correctly (which is usually assumed to mean richly at least).

<sup>45</sup> We harbor no such hopes because we believe representations are not necessary and appear only in the eye or mind of the observer.

<sup>46</sup> Neural networks is the parent discipline of which connectionism is a recent incarnation. Workers in neural networks claim that there is some biological significance to their network nodes, as models of neurons.

Brooks afirma que seu projeto não tem nenhuma pretensão de ser biologicamente mais plausível (BROOKS, 1991a, p. 147).

O objetivo de Brooks, de fato, consistia na construção de sistemas autônomos que reagiriam por si mesmos aos estímulos do meio, de forma satisfatória. Posteriormente, em 2008, Brooks fundou a *Rethink Robotics*, uma empresa cujos objetivos dizem respeito à construção de robôs com uma maior praticidade comercial. O robô Baxter (2012), por exemplo, foi construído a partir da arquitetura de subsunção para realizar trabalhos na produção industrial. Contemporaneamente, podemos apontar como uma tentativa de efetivar o projeto iniciado por Brooks os modelos robóticos construídos baseados na reatividade proporcionada por uma linguagem de programação. Em outras palavras, trata-se de um novo paradigma de linguagem de programação cujo escopo é de tal maneira formalizado que permite aos modelos robóticos terem “reações” aos estímulos do meio (ROBERTSON & WILLIAM 2005). Nesse sentido, as ações e o desempenho do modelo não são totalmente previstos pelo programador, mas dizem respeito à amplitude de seu algoritmo.

Desse modo, a abordagem sugerida por Brooks possibilitaria uma modelagem incremental, ou seja, uma modelagem que teoricamente pode construir desde sistemas simples até sistemas inteligentes autônomos (BROOKS, 1991a). Nas palavras de Brooks (1991a, p. 149) “a ideia aqui é utilizar o mundo como seu próprio modelo [...]”, além do mais o modelo computacional deveria *perceber* constantemente o ambiente, e assim possivelmente ter algum tipo de comportamento inteligente. A inteligência, nesse sentido, diz respeito à apresentação de comportamentos satisfatórios e bem sucedidos em ambientes não controlados. Argumentamos que o processo envolvido na incrementalização dos comportamentos dos modelos pode ser analisado a partir da perspectiva da Teoria da Auto-Organização. Desse modo, podemos analisar os modelos computacionais de, pelo menos, duas perspectivas: em um plano, o modelo pode ser auto-organizado em si, uma vez que não existe um controlador central das camadas reativas que desencadeia os comportamentos do modelo. E em outro plano, as relações estabelecidas entre o modelo computacional de Brooks e o ambiente em que ele está situado podem ser consideradas auto-organizadas, pois não são pré-determinadas e acontecem no momento presente, dadas as condições da construção do modelo e as características do ambiente.

Trataremos na sessão seguinte da possível relação entre o projeto de modelagem computacional de Brooks que destoa das abordagens tradicionais e a Teoria da Auto-Organização. Desse modo, pretendemos apontar para um paradigma de análise que considere um sistema inteligente aquele capaz de aprender e ajustar suas ações por meio de *feedbacks*

resultantes de sua interação com o ambiente. Para tal, em consonância com a proposta de Brooks, argumentaremos que o agente envolvido nestes processos necessita ser autônomo.

### 3.2 A NOVA ROBÓTICA E A AUTO-ORGANIZAÇÃO: UM POSSÍVEL PARADIGMA PARA A MODELAGEM COMPUTACIONAL DA INTELIGÊNCIA?

Na sessão anterior apontamos que o agente envolvido nas abordagens metodológicas, tais como a apresentada por Brooks, necessitaria ser *autônomo*. Mas em que consiste a autonomia e como modelá-la em sistemas artificiais? Para adentrar nesta discussão, primeiramente caracterizaremos as distinções entre sistemas humanos e sistemas artificiais (GONZALEZ et al, no prelo).

Segundo Gonzalez et al. (no prelo), o modelo humano (*soft system*) incorpora certos parâmetros de controle de modo a ter a habilidade de mudar a direcionalidade de suas ações, possuir flexibilidade cognitiva para lidar com contradições, dúvidas e emoções, ter a habilidade de perdoar o passado e tomar decisões baseadas no conteúdo semântico e significativo das informações. Desse modo, são caracterizados como *sistemas abertos* cuja tomada de decisão é desencadeada por *feedbacks* relacionados à história do sistema. Os modelos artificiais, por outro lado, incorporam a aprendizagem de estratégias para comportar-se em situações específicas, tomam decisões baseados na eficiência técnica e possuem certos mecanismos de auto-organização, mas em geral constituem sistemas fechados na medida em que suas partes são passivas e podem ser observadas.

Segundo Gonzalez et al., outra diferença fundamental entre seres humanos e sistemas artificiais consiste em que os primeiros podem ajustar-se, melhorando suas habilidades e consequentemente seu desempenho a partir de experiências próprias. Os sistemas artificiais, no entanto, dado seu fechamento, podem melhorar suas performances, mas apenas a partir da repetição, do hábito e do treino. A aprendizagem humana é autônoma nesse sentido, pois parte do sistema em si, enquanto a *aprendizagem artificial* é condicionada por um conjunto de parâmetros de controle (linguagem de programação, programador, engenheiros e arquitetura física segundo o qual os modelos são construídos) (GONZALEZ et al, no prelo).

O sistema artificial, nesse contexto, é atrelado à finalidade para a qual foi desenvolvido. Finalidade esta que foi imposta exteriormente (pelos programadores, engenheiros e demais responsáveis pela sua construção) e condicionada por uma estrutura física. Nesse sentido, a mudança de comportamento será, na verdade, uma tentativa de atingir da melhor maneira a finalidade já estabelecida. Como apontado por Gonzalez et al:

[...] [o agente artificial] pode mudar suas ações de acordo com as mudanças no ambiente, condições ou recursos, e pode até mesmo "experimentar" de forma aleatória, na tentativa de encontrar ações que sejam mais eficientes. Sistemas futuros podem até ser capazes de programar-se de modo a obter os melhores algoritmos para os fins pretendidos. No entanto, "alguém" tem de definir os efeitos dos sistemas. Máquinas não têm desejos ou necessidades. Eles são construídos para um propósito: para fornecer serviços ou produtos, a fim de satisfazer os desejos ou necessidades humanos<sup>47</sup> (GONZALEZ et al, no prelo).

Desse modo, os sistemas artificiais não teriam um grau forte de autonomia, uma vez que estariam condicionados a propósitos impostos externamente e, nesse sentido, não poderiam ser considerados *agentes* na mesma acepção em que seres humanos, por exemplo, são considerados. Para ser considerado um *agente*, o sistema artificial deveria apresentar autonomia, capacidade de reação e capacidade pró-ativa, tal como caracterizado por Brooks (1991a, 1991b). Assim caracterizada, a noção de *agente* aponta para uma possível definição de autonomia. Nas palavras de Franklin e Graesser (1996, p. 25) “Um agente autônomo é um sistema situado em e fazendo parte de um ambiente, percebendo-o e agindo sobre ele, ao longo do tempo, buscando seus próprios objetivos e percebendo o que deseja para o futuro<sup>48</sup>”. Agentes autônomos<sup>49</sup> são adaptativos, aprendem e tomam decisões. Como aponta *The Royal Academy of Engineering*:

Temos a expectativa de que sistemas autônomos possam simular seres humanos, aprendendo com experiências e tomando decisões? Será que o famoso “Teste de Turing” se aplica aqui? Em algumas circunstâncias, um sistema autônomo [artificial] pode ser mais previsível e confiável do que um ser humano. Um sistema autônomo [artificial] que “toma decisões” será, em grande medida, determinado por entradas e resultados de operações passadas, e não seria “impetuoso” da mesma forma que um ser humano pode ser. Isto é o que os torna tão potencialmente valiosos para tarefas que exigem tomadas de decisões rápidas em circunstâncias potencialmente perigosas. No entanto, em situações complexas, a amplitude da experiência humana pode dar origem a um julgamento melhor do que os produzidos por um sistema

<sup>47</sup> It can change its actions according to changes in the environment, conditions, or resources, and can even “experiment” randomly in the attempt to find actions that are more efficient. Future systems could be able to program themselves in order to obtain the best algorithms for their intended purposes. However, “someone” has to **define** the purposes of the systems. Machines have no desires or needs. They are built for a purpose: to provide services or products in order to satisfy human desires or needs.

<sup>48</sup> An autonomous agent is a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future.

<sup>49</sup> Alguns projetos na IA contemporânea tentam modelar agentes autônomos a partir de um paradigma híbrido. Modelos híbridos se baseiam na combinação de técnicas de arquiteturas reativas e cognitivas, configurando uma alternativa para sanar as deficiências de ambos os paradigmas. Nesse sentido, os modelos cognitivos apresentam incapacidade de reação rápida e adequada a situação não previstas, enquanto os modelos reativos apresentam a incapacidade de estabelecer alternativas comportamentais quando a situação atual diverge do objetivo dado. Sendo assim, os modelos híbridos surgem como uma alternativa na medida em que utilizam planejamento de alto nível de abstração enquanto outras camadas reagem à estímulos do meio. Esse planejamento garante ao modelo maior rapidez de processamento e consequentemente de resposta comportamental (MURPHY, 2000).

"programado" por um conjunto limitado de comportamentos anteriores<sup>50</sup>  
*(The Royal Academy of Engineering, 2009, p. 2)*

Nesse sentido, um sistema autônomo é caracterizado basicamente pela tomada de decisões baseadas em experiências passadas, isso significa que um sistema autônomo é capaz de aprender, ajustar suas ações e responder a situações amplamente. O aprendizado, portanto, pode proporcionar novas e não previstas capacidades ao agente. E a novidade apenas emerge em um contexto auto-organizado, ou seja, no qual não existe a predominância de um agente central e controlador (tal como caracterizamos na sessão anterior deste trabalho).

Como apontado por Teixeira (1998, p. 137), um agente *autônomo* como o construído por Brooks (1991a, 1991b) “é uma coleção de comportamentos competindo entre si. Do caos inicial, um padrão coerente de comportamento vai sendo gerado; há uma aposta na auto-organização do comportamento”. Como já apontamos anteriormente, segundo Debrun:

Há auto-organização cada vez que, a partir de um encontro entre elementos realmente distintos (e não analiticamente) distintos, desenvolve-se uma interação sem supervisor (ou sem supervisor onipotente) – interação essa que leva eventualmente à constituição de uma “forma” ou à reestruturação, por “complexificação”, de uma forma já existente (DEBRUN, 2009, p. 63)

A auto-organização dos comportamentos computacionais é possível porque, ao ser construído, o modelo é testado diretamente no mundo real, em oposição ao mundo puramente virtual em que são testados os modelos computacionais tradicionais. Assim, os modelos desenvolvidos por Brooks poderiam participar de processos auto-organizados na medida em que seu comportamento se dá na interação direta com o meio em que está inserido a partir de suas próprias estruturas. Mais especificamente, os modelos de Brooks poderiam ser considerados como participantes de uma auto-organização secundária, pois partem da organização de um sistema já estabelecido (o próprio modelo).

Podemos analisar os possíveis agentes autônomos artificiais a partir da Teoria da Auto-organização na medida em que “A auto-organização não sendo uma questão de tudo ou nada, mas de mais ou menos, podem intervir – na organização de um ser, de um artefato ou de uma

---

<sup>50</sup> Do we expect autonomous systems to simulate humans, learning from experience and making decisions? Does the famous ‘Turing Test’ for a thinking machine apply here? In some circumstances, an autonomous system may be more predictable and reliable than a human. An autonomous system that ‘makes decisions’ will to a large extent be determined by input and past operation, and would not be ‘impetuous’ in the way that a human could be. This is what makes them so potentially valuable for tasks that require quick decisions in potentially dangerous circumstances. However, in highly complex situations, the breadth of human experience could give rise to better judgements than those made by a system ‘programmed’ by a narrow range of previous behavior.

situação – outros princípios que não a auto-organização, ao lado dela ou em concorrência com ela.” (DEBRUN, 2009, p. 55). Trata-se de graus de auto-organização que se desenvolvem temporalmente.

Nesse contexto, a materialidade do sistema envolvido no processo de auto-organização não é determinante, mas a temporalidade envolvida pode ser. Como apontado ainda por Debrun:

Conforme a estrutura da temporalidade – ou seja, a relação nela existente entre passado, presente e futuro – a natureza ou o grau da auto-organização será diferente. Se, por exemplo, o peso do passado for esmagador – como parece ser o caso na memória biológica, vegetal ou animal -, a eventualidade de uma auto-organização “secundária” (epigênese) será muito limitada no nível de espécie. Se, inversamente, a abertura do futuro (no nível humano) for extrema – sob a forma de projetos de re-auto-organização que pretendam fazer *tabula rasa* do passado -, esses projetos, na falta de âncora, tenderão a fracassar. (DEBRUN, 2009, p. 140-141)

A questão que colocamos aqui é: qual seria o papel da temporalidade nos modelos de Brooks? A estrutura temporal ideal para a auto-organização parece ser aquela que não condiciona o processo, mas serve de alicerce para lançá-lo ao futuro, por assim dizer. Nos modelos computacionais, cuja estrutura é basicamente física, como lidar com o tempo envolvido nos processos? Brooks não aborda explicitamente a temporalidade, mas aponta que a inteligência dos modelos computacionais, construídos segundo a arquitetura de subsunção, está relacionada com a interação com um ambiente cuja complexidade é aumentada progressivamente. Tal aumento proporcionaria ao modelo *aprender* comportamentos cada vez mais sofisticados, baseados em comportamentos anteriores mais simples. Esse processo de incrementalização (como vimos, quando comportamentos considerados simples propiciam comportamentos mais complexos) necessita de *tempo*. Nesse contexto, o tempo parece estar relacionado com o conjunto de comportamentos presentes que pode ser decorrente da complexificação de comportamentos anteriores. Podemos dizer que esse processo envolve um tempo *experienciado*, além daquele presente no processamento físico do modelo (o tempo que as informações levam para perpassar as estruturas eletro-físicas do modelo).

Em suma, o projeto de Brooks diz respeito à construção de modelos computacionais que, devido à sua constituição física e lógica, são capazes de complexificar seus comportamentos tornando-os cada vez mais sofisticados. Esse processo parte da interação entre o modelo e os desafios de um ambiente dinâmico (no sentido de não ser controlável). A partir desses desafios impostos pelo meio, o modelo, dadas as condições iniciais propícias e a metodologia de construção computacional empregada, *reage* e se coloca em um mundo real não controlado.

A proposta de Brooks parece teoricamente promissora, mas na prática as tentativas de construção de robôs autônomos a partir dela não se mostram totalmente satisfatórias. Como aponta Haselager (2004):

Há bastante trabalho promissor nas áreas de robótica reativa e híbrida (Arkin, 1998; Brooks, 1999; Beer, 1997; Mataric, 1998; Werger, 1999; Murphy, 2000), mas é claro que estes robôs estão muito longe de exibir o tipo de comportamento comum que nós, seres humanos, mostramos quando acordamos, fazemos café da manhã e vamos ao trabalho. Mesmo assim, parece que os modelos da robótica reativa são mais apropriados para melhorar nossa compreensão do comportamento habitual. Talvez pudéssemos dizer que, dadas as ‘metas’ dos robôs (autopreservação, evitar a perda total de energia, fazer tarefas simples como encontrar ou guardar objetos), eles fazem aquilo que precisam fazer, na maior parte do tempo. Neste sentido, muitos robôs mostram ‘o seu’ tipo de comportamento comum. Contudo, não podemos esquecer que vai ser muito difícil aumentar este desempenho para atingir a complexidade humana (HASELAGER, 2004, p. 229)

Uma das críticas referentes ao paradigma de arquitetura reativa é a falha de rapidez nas respostas dadas pelos modelos em ambientes não controlados (MURPHY, 2000), pois ainda há a presença, mesmo que não central, de representações. Paradigmas inspirados minimamente na manipulação de representações no processo cognitivo são criticados por autores como Chemero (2009).

Para sanar as falhas deixadas pela manipulação de representações (como a falta de rapidez de processamento), Chemero sugere uma abordagem denominada *Radical Embodied Cognition Science*. Segundo Chemero, a ciência cognitiva incorporada radical pode ser definida como: “... o estudo científico de percepção, cognição e ação como fenômenos necessariamente incorporados, usando ferramentas exploratórias que não postulam representações mentais. É a ciência cognitiva sem ginástica mental” (2009, p. 512). Basicamente, a ciência cognitiva incorporada radical aborda a necessidade do ambiente e o valor da exploração dinâmica nos processos cognitivos, mas combina com esses pressupostos a teoria computacional da mente. Para isso, a ciência cognitiva incorporada radical, segundo Chemero (2009, p. 512), utiliza ferramentas metodológicas assentadas na Teoria dos Sistemas Dinâmicos. Desse modo, os estados cognitivos são considerados propriedades de um agente situado em um meio e estabelecendo relações dinâmicas com ele, sem a mediação constante de representações e manipulações simbólicas.

Nesse sentido, a ciência cognitiva incorporada radical é consoante com os pressupostos da Teoria da Cognição Situada e Incorporada (conforme já caracterizada neste trabalho), mas também está atrelada à teoria computacional da mente e se vale de conceitos da

Teoria dos Sistemas Complexos para caracterizar os processos mentais. Os trabalhos de Randall Beer (2003), por exemplo, se enquadram na vertente da ciência cognitiva incorporada radical.

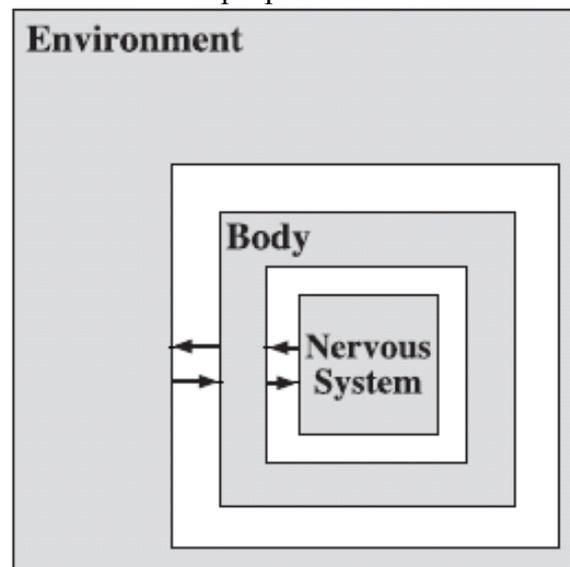
Beer (2003) argumenta que a dinâmica das performances e comportamentos estratégicos pode ser analisada a partir do acoplamento global do sistema cérebro-corpo-ambiente. A interação entre o agente (considerado enquanto corpo, que inclui um cérebro mas não se limita a ele) e o ambiente permite que tal acoplamento seja observado. É o comportamento do agente que revela o tipo de relação estabelecida com o meio e como ambos se influenciam. No que diz respeito à modelagem, Beer propõe um modelo baseado em redes neurais como ferramenta explicativa dos comportamentos desses agentes dinâmicos. Em um primeiro momento, segundo Beer (2003, p. 209): “O corpo físico de um agente, a estrutura de seu ambiente e seu contexto social podem exercer um papel tão importante na geração do seu comportamento quanto seu cérebro. De fato, em um sentido muito real, a cognição não pode mais ser vista como limitada a cabeça de um agente [...]”. Trata-se de um agente incorporado em um meio, munido de experiências passadas que lhe proporcionam um agir da melhor forma possível em um contexto dinâmico.

Nesse sentido, existe uma dimensão dinâmica da cognição na qual os comportamentos do agente surgem da interação constante entre seu cérebro, seu corpo e seu ambiente e, como aponta Beer (2003, p. 210), “Sob esta perspectiva, o foco se desloca da representação precisa de um ambiente para a interação contínua com o ambiente por meio do corpo, estabelecendo padrões de comportamento que são adaptativos para o agente<sup>51</sup>”, tal como representado na figura 6. Nessa perspectiva, o agente não atua no mundo mediado por representações, mas age diretamente no meio em que está situado. Para tentar modelar computacionalmente essa concepção, Beer baseia sua proposta em estados neuronais e em ferramentas matemáticas que podem ser utilizadas para caracterizar a estrutura de um espaço possível para a trajetória de comportamentos dinâmicos.

---

<sup>51</sup> On this view, the focus shifts from accurately representing an environment to continuously engaging that environment with a body so as to stabilize coordinated patterns of behavior that are adaptive for the agent.

**Figura 6** - Diagrama explicativo da proposta de Beer.



**Fonte:** Beer (2003, p. 211).

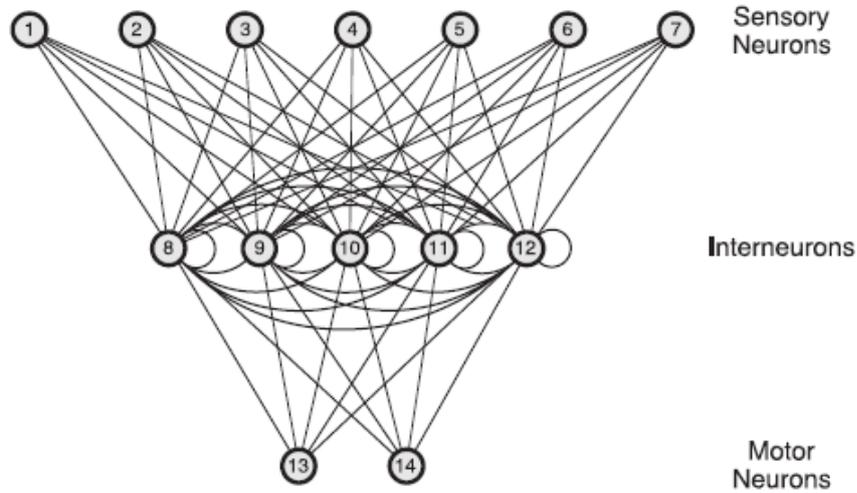
Uma abordagem dinâmica dos comportamentos coloca em questão o papel necessário das representações e da computação nas teorias cognitivas (BEER, 2003, p. 210). Qual seria o papel das representações em sistemas dinâmicos como as redes neurais? E essa dinâmica seria passível de modelagem computacional?

Tradicionalmente, nos modelos conexionistas baseados em redes neurais, os elementos básicos de análise são os *neurônio-símile*, unidades cujo objetivo no modelo é serem correlatas aos neurônios biológicos. Para tanto, os neurônios-símile possuem níveis de ativação que variam de acordo com os estímulos recebidos do meio externo. Ao serem ativados, os neurônios-símile emitem sinais a outros neurônios e assim acabam formando uma rede estruturada. Tal rede se constitui de uma camada formada de unidades de entrada (*inputs units*), uma camada formada de unidades de saída (*output units*) e uma camada intermediária (*hidden units*) que realiza a comunicação entre a camada de entrada e a de saída. A rede neural artificial funciona de acordo com os estímulos recebidos do meio externo. Sua principal função é desempenhar a tarefa que os neurônios desempenham no cérebro para o desencadeamento de um processo cognitivo. Sendo assim, as camadas das redes neurais artificiais desempenham a função de processamento de representações de informações *regulares* advindas do meio ambiente externo. Essa representação é realizada através de pesos das conexões sinápticas entre as unidades neurônio-símiles. Em um primeiro momento o peso das conexões é arbitrário, por meio de treino e regras empregadas de acordo com o estímulo externo, a rede pode se auto ajustar.

Uma vez realizadas as representações dos padrões informacionais do meio externo, a rede neural artificial é capaz de desenvolver generalizações e, assim, realizar antecipações e reconhecimentos de padrões ambientais. Em suma, o processamento das redes neurais artificiais consiste em: 1) representar, a partir de pesos e regras, as informações presentes no ambiente; 2) apreender os padrões informacionais; 3) realizar generalizações a partir dos padrões apreendidos e 4) ser capaz de antecipações não contidas anteriormente na rede.

A proposta de Beer consiste no desenvolvimento de *redes neurais de tempo contínuo*. Desse modo, não haveria um tempo específico para a representação de pesos e regras, sendo o aprendizado uma propriedade emergente das relações estabelecidas entre a gente e meio no tempo presente. Dessas relações, segundo Beer, emergem *comportamentos cognitivos minimalistas*, necessários para o comportamento complexo. Os comportamentos são minimalistas na medida em que emergem da relação dinâmica entre agente e mundo. Desse modo, os comportamentos do agente dizem respeito ao ajuste de uma situação presente. Tal ajuste sana a “perturbação” presente, mas também passa a constituir o conjunto de comportamentos possíveis desse agente. Em uma situação mais complexa, por exemplo, esse comportamento *minimalista* pode integrar outro conjunto de comportamentos de modo a propiciar ao agente um ajuste mais complexo. A figura 7 mostra as relações que o agente concebido por Beer estabelece. O projeto de Beer, nesse sentido, é baseado na proposta das redes neurais no qual existe um neurônio sensorial que *capta* as informações do meio, em seguida essa informação passa para uma camada intermediária que transfere a informação relevante para os neurônios motores que desencadearam um dado comportamento. A proposta de Beer destoa nas redes neurais tradicionais na medida em trabalha operando um *tempo contínuo*, ou seja, a rede não para de processar as informações captadas pelos neurônios sensoriais. Isso garante um processamento de informações rápido que acarreta na diminuição do hiato entre o estímulo recebido e a resposta dada. A relação entre o *input* e o *output* é quase que imediata.

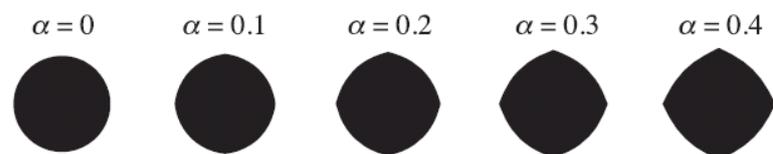
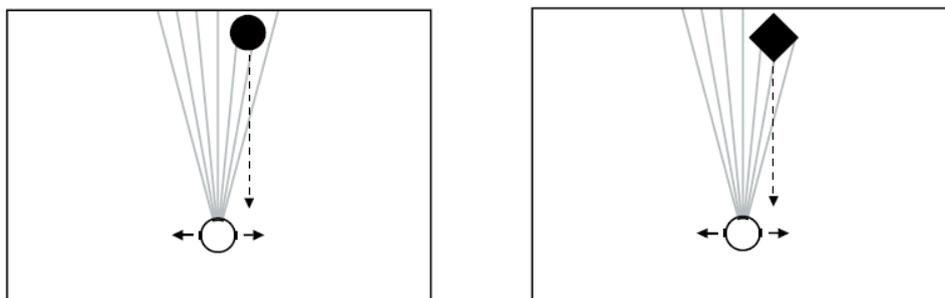
**Figura 7** - Diagrama das relações estabelecidas entre o agente e o meio em que ele está situado a partir da mediação de redes neurais.



Fonte: Beer (2003, p. 213).

Todavia, o comportamento do agente é mediada por relações matemáticas, embora Beer afirme que os comportamentos sejam minimalistas e não se assentem em uma representação interna e central. O movimento do agente no espaço, por exemplo, se dá a partir de um cálculo matemático. Tal como aponta Beer (2003, p. 212-213), o agente reconhece (pois as diretrizes de ação já são dadas previamente em sua programação) um padrão de forma de um objeto qualquer do meio, calcula seu tamanho e, a partir do resultado, calcula seus próprios movimentos e velocidade de acordo com certos objetivos.

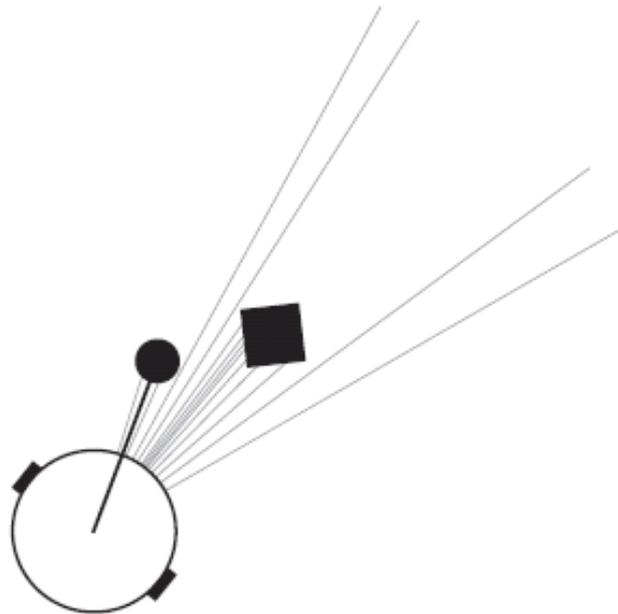
**Figura 8** - Reconhecimento de padrões segundo a proposta de Beer.



Fonte: Beer (2003, p. 213).

A figura 8 ilustra esse processo de reconhecimento. O agente detecta um objeto no meio, reconhece sua forma (se ele é circular ou tem forma de diamante), calcula o tamanho desse objeto e a partir disso traça sua rota. Por exemplo, o agente detectou um balde no meio do caminho até chegar a uma porta de saída. O agente necessita sair pela porta, então ele: 1) reconhece o obstáculo, sua forma, seu tamanho; 2) pelos cálculos de que o agente é dotado previamente, o resultado foi que o objeto em questão é circular e tem o raio de 0.1 e 3) o agente agora sabe que seu obstáculo possui o raio de 0.1, então ele se desloca 0.1 em uma velocidade propícia para contornar o obstáculo. Para manipular o meio em que está inserido, o agente possui, além de um aparato perceptual, dois motores (para a locomoção) e um braço móvel, como vemos na figura 9. Nota-se aqui a relevância de algo que se assemelhe a um *corpo*, ou que pelo menos, desempenhe funções análogas às desempenhadas por ele.

**Figura 9** – Ilustração do agente guiado por percepções, segundo Beer. O agente é o círculo maior e possui um olho que abarca um campo de visão (linhas cinza), dois motores (os retângulos escuros nas laterais do círculo maior) e um braço (linha escura).



**Fonte:** Maes *et al.* (1996, p. 422).

Os cálculos realizados podem ser considerados representações que fazem a mediação dos comportamentos do agente com o ambiente. Além do mais, existe a estrutura física do qual o agente é constituído e que também pode ser considerada uma mediação entre a informação captada e o comportamento resultante. Em um primeiro momento parece existir uma tentativa de matematizar comportamentos comuns cognitivamente carregados, mas,

segundo Beer (1996, p. 422), é o agente que deve ser simples o suficiente para ser tratado matematicamente e computacionalmente de modo que ele possa “evoluir” utilizando técnicas que, no máximo, incrementarão certos comportamentos. Apontamos aqui para o limite dessa abordagem: Não seria o projeto de Beer uma tentativa de matematizar comportamentos decorrentes de habilidades cognitivamente carregadas (como andar, superar obstáculos, mover objetos, etc)? Em que esta abordagem contemporânea difere dos projetos dos pioneiros da IA funcionalista? Reconhecemos que Beer agrega conceitos que enriquecem a discussão acerca da modelagem de sistemas complexos, mas ainda se pauta em um dos pilares mais fortes da IA tradicional que busca a matematização de comportamentos naturais a fim de simulá-los mecanicamente.

Beer propõe um paradigma de modelagem de comportamentos baseado na ausência de representações internas e assentado *na incrementalização* destes. Apontamos aqui que, os comportamentos podem sim ser considerados como mediados por representações, mesmo que estas não ocupem um lugar central no processo de suposta aquisição e desenvolvimento destes. Nesse sentido, argumentamos que os cálculos matemáticos envolvidos no comportamento do agente representam o modo como tais comportamentos se dão no plano da ação. São, portanto, uma representação de uma característica natural e constituem uma mediação entre a informação recebida e o comportamento do agente artificial.

Todavia, Beer argumenta que da relação dinâmica entre agente e meio emergiriam comportamentos satisfatórios para o contexto. O conceito de *emergência* será mais bem definido na sessão seguinte, mas para o presente momento podemos considerar que um comportamento é *emergente* se ele não pode ser totalmente reduzido às partes que o propiciaram. Segundo Beer, como já apontado anteriormente, os comportamentos do agente são resultantes da sua relação com o meio. Nesse sentido, embora o agente seja munido de complexos cálculos matemáticos, estes de nada serviriam se o agente não estivesse situado em um meio que proporciona a utilização desses cálculos para a ação. Nessa perspectiva, o comportamento do agente proposto por Beer é emergente, pois não se reduz aos cálculos, embora seja intrinsecamente dependente deles.

A proposta de Beer se assemelha ao paradigma incremental e reativo proposto por Brooks, pois ambos tratam da construção de modelos computacionais situados e incorporados estabelecendo relações dinâmicas com o meio, mantendo um papel não central para as representações. No entanto, Beer (1996, p. 428) ressalta que o projeto de Brooks (incorporado no robô Cog) não se assenta em pressupostos das redes neurais ou na ideia de sistemas

dinâmicos. Para Beer (1996, p. 428) os objetivos do projeto Cog são muito mais ambiciosos ao utilizar um robô físico como ferramenta de trabalho.

A novidade é que Beer considera os agentes situados e incorporados como participantes de relações dinâmicas. A dinâmica das relações, nesse sentido, proporciona aos agentes a capacidade de complexificar seus comportamentos. Seriam, então, os agentes propostos por Brooks e Beer considerados sistemas complexos de fato? Como se daria esse processo? Na sessão seguinte abordaremos questões deste tipo.

### 3.3 HABILIDADES COMO UMA CAPACIDADE DE SISTEMAS COMPLEXOS: LIMITES E POSSIBILIDADES DA MODELAGEM COMPUTACIONAL DA NOVA ROBÓTICA

Como apontado por Bresciani e D'Ottaviano (2000, p. 283), o ramo da sistêmica remonta pelo menos à década de 50 com estudos desenvolvidos sob a denominação de *Teoria Geral de Sistemas*. Tais estudos pretendem analisar, de forma interdisciplinar, a organização abstrata dos fenômenos, independentemente de sua constituição física ou configuração atual. A sistêmica, portanto, propõe investigar os *princípios gerais* que regem a organização e desenvolvimento dos fenômenos. Nesse sentido, situamos o presente trabalho sob este arcabouço teórico por julgar que aqui encontramos ferramentas adequadas para a análise de sistemas artificiais, sua relação com informação e complexidade e sua eventual capacidade de instanciar habilidades.

Assim sendo, o primeiro passo deste trabalho será apresentar o conceito de *sistema* cuja definição é basilar para nosso propósito. Como apontado por Bresciani & D'Ottaviano:

Um sistema pode ser inicialmente definido como uma entidade unitária, de natureza complexa e organizada, construída por um conjunto não vazio de elementos ativos que mantêm relações, com características de invariância no tempo que lhe garantem sua própria identidade. Nesse sentido, um sistema consiste num conjunto de elementos que formam uma estrutura, a qual possui uma funcionalidade (BRESCIANI & D'OTTAVIANO, 2000, p. 284-285).

Tal como elucidado na definição acima, um sistema é uma entidade complexa e organizada constituída de elementos que interagem entre si, proporcionando, assim, a emergência de uma funcionalidade. Nota-se que, sob a perspectiva da sistêmica, não existe menção à constituição física dos elementos ou da estrutura por eles constituída. Sendo assim,

este plano de análise abarca sistemas orgânicos e artificiais, pois está interessado nos *princípios gerais de organização*.

Sob a perspectiva sistêmica, a análise do sistema é sempre uma construção no sentido de que tal sistema é “[...] concebido pelo sujeito, que também pode lhe atribuir finalidade.” (BRESCIANI & D’OTTAVIANO, 2000, p. 284). Em outras palavras, a análise de um sistema está atrelada ao olhar de um sujeito que estabelece uma espécie de recorte da realidade e delimita elementos, estruturas, funções e funcionalidades. Vale notar que o sujeito da análise sistêmica está situado em um ambiente e em um tempo e seu plano de análise se relaciona com certos objetivos específicos. O plano de análise diz respeito tanto ao sistema em si, quanto aos elementos que o constituem. Os elementos do sistema, por sua vez:

[...] são considerados como sendo as partes, os componentes, os atores ou os agentes que realizam atividades (bem como ações, reações, retroações, proações e transações), conduzem processos e operações, produzem fenômenos e são responsáveis por transformações, conversões e eventos que ***caracterizam os seus comportamentos***. (BRESCIANI & D’OTTAVIANO, 2000, p. 285, destaque nosso).

Nesse sentido, os elementos de um sistema podem ser tanto células e tecidos quanto engrenagens e regras lógicas de programação. Sob o olhar da sistêmica, a natureza dos elementos é variada, mas desempenhará um papel importante na construção da funcionalidade do sistema. Elementos se relacionarão de uma forma específica e constituirão estruturas que desencadearão certas funcionalidades, sendo assim, desempenham um papel causal no comportamento do sistema. No entanto, o processo sistêmico não é de natureza puramente determinística: a funcionalidade do sistema procede da estrutura formada da relação estabelecida pelos elementos temporalmente. Tal relação é constituída no plano da ação por meio da comunhão de parâmetros estruturais (dos próprios elementos e do ambiente) e temporais, cujo produto final não pode ser totalmente previsto dadas as variáveis iniciais.

O tipo de relação estabelecida entre os elementos constituintes do sistema e destes com o meio ambiente, pode caracterizar o processo como complexo. Em um sistema complexo existe a interdependência entre elementos que podem estabelecer relações de diferentes tipos. Como apontado por Bresciani e D’Ottaviano, tais relações podem ser assim classificadas:

[...] relações de primeiro tipo são aquelas que são funcionalmente necessárias e que, portanto, caracterizam o sistema; relações do segundo tipo são aquelas complementares, que acrescentam algo ao comportamento do sistema e que são denominadas relações sinérgicas positivas ou de

cooperação; relações de terceiro tipo são relações que indicam contradições antagônicas, discordantes e de oposição entre os elementos e que são denominadas relações sinérgicas negativas ou de competição; as relações de quarto tipo são aquelas que duplicam as relações existentes (relações redundantes). (BRESCIANI & D’OTTAVIANO, 200, p. 285).

As relações presentes em um sistema podem ser de mais de um tipo e podem mudar de configuração ao longo do tempo. Mas, é a presença dessas relações que possibilita a emergência de propriedades e, conseqüentemente, a evolução do sistema. As propriedades do sistema que podem ser consideradas como emergentes são aquelas que não estavam presentes no sistema de fato, mas que potencialmente se encontravam, como latentes, em seus elementos e, conseqüentemente, em suas estruturas. A capacidade humana de pressionar botões, por exemplo, pode ser considerada uma propriedade emergente de um sistema (o próprio ser humano). Esta capacidade é decorrente da anatomia biológica (polegares opositores, músculos e estrutura cerebral) e do contexto (a tecnologia que permite que sejam construídos artefatos com botões). Trata-se de uma relação circular na qual o contexto oferece condições para a emergência de uma propriedade potencialmente presente em uma determinada estrutura e esta, por sua vez, reestrutura o contexto. Um sistema complexo necessita conter pelo menos uma relação circular, assim há uma sensibilidade às condições iniciais de modo que uma mudança nas relações entre os sistemas pode resultar em uma mudança no comportamento global do sistema.

No caso de sistemas artificiais, propriedades emergentes são mais dificilmente apontadas, pois podemos considerar que todas as propriedades do sistema foram previamente determinadas pela engenharia. Todavia, projetos contemporâneos colocam esta afirmação em xeque (como BROOKS, 1991a). A ideia é que sistemas artificiais, dado o modo como lidam com os dados informacionais, sejam capazes de transpor sua arquitetura física e complexificar seus comportamentos. Tal transformação ocorre, inicialmente, tendo como base os parâmetros instituídos pelos desenvolvedores mas, ao contrário dos sistemas artificiais clássicos, tais parâmetros não se colocam como *limitadores*, mas sim como *propiciadores* de novos comportamentos, ou seja, de comportamentos emergentes.

Para que possam emergir propriedades, tanto de sistemas artificiais quanto de orgânicos, é preciso que o sistema em questão possua anteriormente a propriedade da *globalidade* e da *possibilidade de novidade*. Como apontado por Bresciani e D’Ottaviano (2000, p. 286): “[...] outras propriedades fundamentais [do sistema] são a globalidade (constituição da unidade global com sua invariância) e a possibilidade de novidade (como, por exemplo, no caso extremo, a constituição da própria existência)”. Tais propriedades precisam

estar, de alguma maneira, presentes (mesmo que de forma latente) nos elementos constituintes do sistema e que posteriormente, estabelecendo relações, constituirão as estruturas funcionais. Nesse contexto, a propriedade da globalidade permite que os elementos constituintes do sistema se relacionem de modo dinamicamente equilibrado no fluxo do espaço e do tempo, de modo a possibilitar a emergência de uma *identidade* do sistema. Já a propriedade da possibilidade da novidade diz respeito à abertura do sistema e permite que, mesmo com sua identidade inerente, o sistema possa transformar-se e complexificar-se.

As propriedades da globalidade e da possibilidade da novidade existem na medida em que os elementos constituintes do sistema estabelecem redes de relações arborescentes e circulares. Tais relações propiciam a comunicação entre os elementos do próprio sistema e destes com o ambiente em que estão inseridos de modo que seja possível a complexificação do sistema a partir de uma unidade global. A comunicação entre as partes propicia a incorporação de novas informações, tornando assim o sistema cada vez mais complexo, pois é dessa maneira que o sistema se adapta, modifica ou exclui comportamentos acarretando um ajuste cada vez mais condizente com o meio. As relações circulares permitem a autoregulação, ou seja, a capacidade de ajuste a partir da troca de informações entre o sistema e o meio que apenas existe dado o grau de autonomia e espontaneidade que os elementos do sistema possuem. Como apontado por Bresciani e D'Ottaviano (2000, p. 302): “A atividade espontânea decorre da existência de grau mínimo de autonomia aos elementos atuantes. Por sua vez, processos recorrentes precisam estar presentes para que os elementos autônomos, em suas atividades, se integrem em uma organização com auto-referência”. Os processos recorrentes são possíveis em relações circulares-arborescentes que permitem o surgimento da evolução enquanto uma criação de novidade. Relações do tipo circular, portanto, podem ser entendidas como indicação de abertura do sistema.

Sistemas abertos, devido ao fluxo informacional recorrente, podem estabelecer relações de tal modo que exista um tipo de *criação*. Em relação aos processos de criação Bresciani e D'Ottaviano assinalam:

A criação decorre da influência de diferentes fatores, particularmente aqueles relacionados aos graus de autonomia e à natureza constitutiva dos elementos do sistema (eventualmente de fronteira), como elasticidade e plasticidade e, e alguns casos, a capacidade de imaginação e de concepção. Mas, é importante também considerar a influência dos fatores relacionados à existência de uma organização no sistema propícia às transformações ou à existência de um meio ambiente motivador (incentivador, catalisador e perturbador) do processo de criação. (BRESCIANI & D'OTTAVIANO, 2000, p. 303)

A criação, nesse sentido, está atrelada a processos dinâmicos da relação do sistema no âmbito das partes e destas com o ambiente. É nesse tipo de relação que pode ser garantida a sobrevivência, manutenção e/ou a evolução do sistema. A relação criativa é dinâmica porque envolve sequências de estado de equilíbrio e desequilíbrio que surgem ao longo da transformação do sistema devido à ação de elementos internos e externos. Como apontado, é a dinâmica das relações que favorece a criação de processos, a evolução do sistema e sua manutenção. Nas relações dinâmicas também estão envolvidas as características da regulação e da adaptação, igualmente relevantes para o desenvolvimento do sistema. Como apontado por Bresciani & D'Ottaviano:

Duas características dos sistemas, que estão associadas à manutenção ou à mudança de estados podem ser mencionadas:

- a) Característica da regulação, que se manifesta pela manutenção do estado de equilíbrio e da existência do sistema, frente às contingências internas e externas;
- b) Característica de adaptação, que se expressa pela mudança de um estado em um novo estado de equilíbrio e garante a manutenção da existência do sistema frente às contingências internas e externas. (BRESCIANI & D'OTTAVIANO, 2000, p. 299-300)

Regulação e adaptação estão intimamente relacionadas, pois podemos considerar que a adaptação decorre da regulação do sistema. A regulação, por sua vez, se dá por meio de relações circulares tendo em vista que é nesse tipo de relação que a comunicação (o fluxo de informação em diferentes níveis) entre os elementos e destes com o ambiente é mais propícia. Como apontado por Bresciani & D'Ottaviano (2000, p. 300): “Os mecanismos de regulação e de adaptação, que não são excludentes, surgem das relações dinâmicas no sistema e do sistema com o seu meio ambiente; através deles o sistema mantém sua existência em equilíbrio com o meio ambiente”. É a dinamicidade do sistema que permite que ele estabeleça relações circulares e possivelmente se autorregule e se adapte às condições externas. Trata-se de constantes relações de *feedback* entre elementos, sistema e ambiente dada a maneira como estes componentes se constituíram ao longo do tempo. O tempo aqui não diz respeito ao tempo cronológico que pode ser medido e cronometrado, mas, sim, ao tempo vivido e experienciado no exercício de atividades. A experiência é característica tipicamente atribuída à sistemas orgânicos-naturais, mas tentaremos contrapor esta ideia no ponto 3. Ao descolar a análise dos elementos para o plano do sistema, no qual o comportamento está atrelado às relações de comunicação informacional, podemos considerar sistemas artificiais como experimentadores possíveis e capazes.

Em todo caso, processos de regulação e adaptação se dão por meio de atividades de elementos internos e externos do sistema. No caso de atividades de elementos internos podemos dizer que os mecanismos de controle são de *auto-regulação* ou de *auto-adaptação*. Um sistema deve ser capaz de regular e adaptar-se, garantindo assim a sua existência e identidade (BRESCIANI & D’OTTAVIANO, 2000, p. 300) sendo assim, deve ser capaz também de corrigir os desvios e compensar os desequilíbrios que possam perturbá-lo causando seu desaparecimento. Nesse sentido, o sistema se encontra em uma situação circular de equilíbrio-desequilíbrio, como apontam Bresciani & D’Ottaviano (2000, p. 300): “Os mecanismos de regulação ou de adaptação fazem parte dos processos que levam o sistema a atingir a condição de equilíbrio ou de desequilíbrio, com manutenção de estado ou com mudança de estado”. A sucessão de estados é o que abre margem à evolução do sistema, uma vez este sendo capaz de se adaptar.

Em suma, procuramos argumentar que a sistêmica, enquanto ferramenta de análise dos processos gerais de organização e funcionamento dos fenômenos, nos oferece instrumentos consideráveis no estudo da complexidade, inclusive, em sistemas artificiais. Sob esta perspectiva, consideramos um sistema como uma estrutura funcional que está inserida em um fluxo de relações dinâmicas que permitem a complexidade do sistema. Apontamos também que, para isso, o sistema deve possuir propriedades como a regulação e adaptação, garantindo plasticidade de comportamento. Esse plano de acontecimentos do sistema é permeado por relações de comunicação e consideramos que tais relações sejam fundadas em *informação*.

Sob esta perspectiva, podemos analisar a possibilidade de complexidade em sistemas artificiais. Ao considerarmos a complexidade enquanto um processo baseado em relações de comunicações (entre elementos de um sistema, entre sistemas e meio ambiente), a análise se desloca para além do substrato no qual os elementos e sistemas são constituídos. Desse modo, a comunicação não está atrelada especificamente à natureza material dos sistemas, mas, sim, à sua natureza *funcional*. Para que haja comunicação entre as partes, o que é relevante é a troca e o fluxo de *informação*, independentemente da natureza do substrato físico que a instancia. Na próxima sessão aprofundaremos a questão da informação nos processos de comunicação e, conseqüentemente, de complexificação de sistemas.

### 3.4 O PAPEL DA INFORMAÇÃO NOS PROCESSOS DE COMUNICAÇÃO E COMPLEXIFICAÇÃO

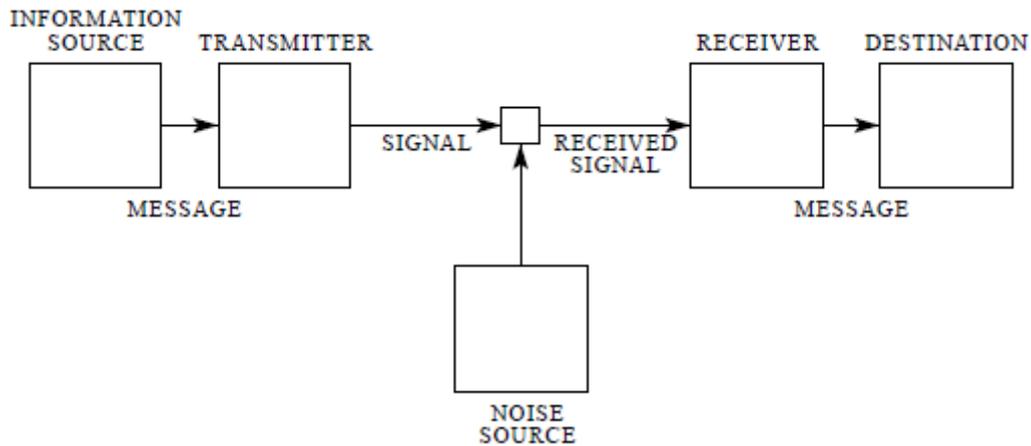
Apontamos a relevância do fluxo informacional no processo de complexificação assentado em processos de comunicação. Argumentamos que a comunicação entre as partes, sistema e ambiente, através do fluxo informacional, possibilita a complexificação do sistema. Nesta sessão, procuramos expor os planos de análise em que o conceito de *informação* é tratado, bem como sua possível relação com a complexidade.

Tradicionalmente, como apontam Gonzalez et al (2004) os estudos referentes ao conceito de informação se ramificam em três planos de análise: 1) plano metodológico; 2) plano epistemológico e 3) plano ontológico. O plano metodológico focaliza a transmissão de mensagem de uma fonte para um receptor, independentemente do significado desta mensagem. Já o plano epistemológico se volta à caracterização da informação e sua relação com o significado. Por fim, os teóricos do plano ontológico estão interessados na natureza da informação.

Tendo em vista que nossa hipótese de que os processos de complexificação de sistemas se dão por meio de relações de comunicação de informação entre as partes, sistema e meio, acreditamos que o plano de análise metodológico da informação pode contribuir na medida em que visa a transmissão de mensagens (considerada um tipo de comunicação). De outro lado, o plano epistemológico da análise da informação é relevante para este trabalho, pois, embora a comunicação entre elementos, sistema e ambiente possa ser descrita como transmissão de mensagens, o *significado* destes é requerido. O processo de comunicação envolvido na complexidade não se resume apenas à transmissão de mensagens, mas envolve também o significado destas mensagens. O processo de complexificação dos sistemas requer transmissão de informação significativa para que, assim, o sistema tenha certa *direcionalidade* e possa adequar seus comportamentos e ações à adaptação e regulagem com o meio em que está inserido.

Podemos apontar como exemplo do plano de análise metodológico da informação os trabalhos de Shannon. O matemático, no texto *A Mathematical Theory of Communication*, traça uma caracterização técnica do conceito de informação já que o seu objetivo é estabelecer um meio de transmissão integral de mensagens entre fonte e receptor. O diagrama abaixo ilustra a teoria de Shannon (1948).

**Figura 10** - Diagrama de transmissão de mensagem segundo Shannon.



**Fonte:** Shannon (1948, p. 380).

A Teoria Matemática da Comunicação envolve os conceitos de fonte (geradora de informação), transmissor (mecanismo que conduz a informação contida nas mensagens até o receptor) e receptor (potencial, que decodifica a mensagem), além dos conceitos subsequentes de sinal, ruído e fluxo informacional. Shannon não trabalha estes últimos conceitos, pois visa um conceito técnico de informação de modo a estabelecer equações que permitam quantificar sua presença (MORAES, 2014, p. 25). Segundo Shannon (1948), a transmissão de mensagens ocorre de uma fonte para um receptor, através de um transmissor, com a mínima intervenção de ruídos, em outras palavras, de elementos que não estavam presentes na fonte original, mas que de algum modo estão presentes no receptor potencialmente final. Nesse contexto, a informação pode ser medida e quantificada objetivamente. Para isso, Shannon propõe a análise binária para a redução do ruído na transmissão de mensagens.

O método proposto por Shannon, assentado na análise binária, é mais econômico, pois apenas trabalha com duas possibilidades (por exemplo, a face de uma moeda ou é cara ou é coroa). Assim, quanto menor for o grau de incerteza (de ruído) contido na mensagem, mais informativa ela é. Podemos ressaltar que a teoria de Shannon é assentada na relação comunicação-informação, no qual a comunicação é necessária para que exista informação. Se aplicarmos a teoria de Shannon na análise de modelos computacionais, podemos ressaltar que, nesse sentido, máquinas<sup>52</sup> podem ser consideradas até superiores a sistemas orgânicos no que diz respeito à transmissão de mensagens e ao aspecto quantitativo do processamento de informação. Mecanismos artificiais são capazes de, em curtos períodos de tempo, transmitir

<sup>52</sup> Neste trabalho utilizamos o termo *máquina* e não sistema artificial quando se trata de apresentar uma teoria específica no qual o autor utiliza o termo *máquina*. Todavia, quando se lê ambos os termos o relevante é o fato de que ambos não são naturais, no sentido de são produzidos a partir de uma técnica e não pela natureza.

um grande número mensagens de uma fonte a receptores e destes para outras fontes, até que um *feedback* adequado à função pré-estabelecida seja desencadeado. O processamento artificial depende da maneira como as máquinas são desenvolvidas. Todavia, para que a comunicação e a quantidade de informação presentes em um sistema sejam relevantes, é necessário que possuam certo tipo de *significado* e assim influenciarem na complexidade de comportamentos e ações do sistema.

O significado da informação pode ser analisado a partir da perspectiva ontológica. Nesse contexto, podemos apontar duas principais vertentes filosóficas que abordam a natureza ontológica da informação: uma *informacional-representacionista* e outra *não-representacionista*. Neste trabalho apresentaremos brevemente a abordagem representacionistas, a partir dos estudos de Frederick Dretske, por julgar ser mais relevante para o objetivo aqui estipulado.

No que diz respeito à vertente informacional-representacionista, a informação existe objetivamente no mundo enquanto um conjunto de regularidades que independem de uma mente externa. Nessa perspectiva, a informação é naturalizada e considerada enquanto uma ferramenta para fundamentar o conhecimento. Para Dretske, os sinais estão puros de informação e carregam diversos *pedaços* desta. A função da mente, nesse contexto, seria reconhecer objetos e eventos no mundo, identificando suas propriedades informacionais (GONZALEZ et al, 2010). Segundo Dretske, isso é possível na medida em que existe um fluxo informacional entre fonte (mundo) e receptor (sujeitos).

Um fluxo informacional se estabelece a partir de relações nômicas (de um para um) nas quais a quantidade de informação presente no receptor é a mesma que é gerada na fonte (o que está no receptor *B*, estava igualmente na fonte *A*). Esse princípio preservaria o mínimo de informação necessária para a relação informacional entre o sujeito e o conteúdo informacional presente no mundo. A partir do princípio que garante a passagem de informação da fonte para o receptor, Dretske fundamenta sua abordagem informacional do conhecimento.

Para Dretske, o conhecimento é um produto que emerge da representação da informação que é percebida no mundo. Assim, Dretske contribui para a solução do problema deixado no *Teeteto* de Platão. O conhecimento, nesse momento, passa a ser considerado como *crença verdadeira justificada em informação*. Nesse contexto, o conhecimento é originário da informação disponível no meio e se efetiva a partir de uma representação de um sujeito com essa capacidade.

Para Dretske, a informação existe no meio objetivamente enquanto elemento natural constituinte do mundo, mas o significado não. Segundo ele, o significado depende de um

sujeito com a capacidade de representar a partir da informação natural. Como apontado por Haselager (2004, p. 106) “As representações são postuladas para atuar como intermediárias entre a percepção e a ação, especificando informações sobre o mundo, frequentemente de forma independente do contexto – e da ação”. Em outras palavras, a representação tomaria o lugar de elementos externos (carregando seus conteúdos), faria a intermediação entre estes e o sujeito cognoscente e, por fim, desencadearia um dado comportamento (HASELAGER, 2004).

Aqui cabe a distinção, estabelecida por Dretske, entre *conteúdo informacional* e *conteúdo semântico*. O conteúdo informacional pode ser entendido como aquilo que o signo indica através de uma relação de dependência necessária. Por outro lado, o conteúdo semântico envolve mecanismos de representações que são desenvolvidas por aprendizagem que se dá em uma determinada temporalidade. Desse modo, o significado é considerado uma propriedade emergente que depende da representação e da aprendizagem. Esta, por sua vez, envolve processos de ajustes e correções a partir de ações de sujeitos situados. Por fim, de acordo com os defensores da vertente informacional-representacionista apenas organismos com capacidades de representar o mundo (tais como os seres humanos) são capazes de lidar com informação significativa, uma vez que graças às representações criadas em suas mentes conseguem corrigir/ajustar suas ações.

Assim, objetivamos tecer considerações acerca da possibilidade de sistemas artificiais serem capazes de se complexificarem, ou seja, de apresentarem um enriquecimento de propriedades, comportamentos e configuração não existentes em seu estado inicial. Os sistemas artificiais poderiam, assim, ser utilizados como instrumentos metodológicos e experimentos de comprovação de teorias. Em outras palavras, os sistemas artificiais, uma vez que são construídos a partir de conjuntos de pressupostos, dependendo de suas performances, poderiam indicar *como* os processos envolvidos na complexidade ocorrem e a que âmbito (natural e/ou artificial) tais processos estão, ou não, limitados, configurando assim uma *ferramenta explicativa*.

Levantamos a hipótese, na esteira de Bresciani & D’Ottaviano (2000), de que a complexidade envolve relações dinâmicas por meio das quais o sistema estabelece comunicação entre suas partes e o meio, acarretando plasticidade de comportamento, o que propicia a complexificação do sistema. Tais relações comunicativas podem desencadear processos de complexificação no sistema porque nelas estão presentes informações significativas que norteiam a adaptação e a possível evolução do sistema. Muitos fatores estão envolvidos em processos complexos e sua modelagem encerra grandes dificuldades técnicas

de engenharia (programação e arquitetura física, por exemplo) e de função (determinar como um sistema deve se comportar para seja considerado complexo). Todavia, existem projetos de pesquisa que visam a construção de máquinas que, dentre outras características, são incorporadas e situadas, capazes de aprender e que contam com um aperfeiçoado sistema de comunicação interna e externa.

Nesse contexto, retomaremos os trabalhos já apresentados de Nobert Wiener e de Rodney Brooks como recurso metodológico de análise da possibilidade de sistemas artificiais complexos. Assim, Wiener em *Cybernetics* apresenta considerações acerca da capacidade das máquinas aprenderem e de se comunicarem. Nesse sentido, Wiener elucida a noção de *aprendizagem* por meio de sistemas artificiais que desempenham a função de *jogar*, no caso o xadrez, como já apresentamos neste trabalho. Nas palavras de Wiener:

Não é muito difícil de fazer máquinas que jogam xadrez de um tipo. A mera obediência às leis do jogo, de modo que apenas movimentos legais são feitos, é facilmente instanciado em computadores bastante simples. De fato não é difícil adaptar uma máquina digital para estes fins. (WIENER, 1965, p. 171)

A facilidade de modelagem do jogo de xadrez apontada por Wiener se justifica pelo fato de que as variáveis envolvidas no jogo são passíveis de redução a termos numéricos (WIENER, 1965, p. 172). Nesse sentido, a definição das regras e do objetivo do jogo seriam o suficiente para afirmar que a máquina realmente estaria jogando xadrez. Além do mais, a máquina conta com uma memória na qual são armazenadas as partidas já finalizadas, que servirão de base para jogos futuros. A máquina, então, segundo as informações pré-estabelecidas acerca das regras e do objetivo do jogo, analisa as partidas realizadas abstraindo dados relevantes delas, como as possibilidades de movimentação das peças. Desse modo, a máquina desenvolve a sua “personalidade” enquanto jogador e *aprende*, mesmo em um sentido bastante restrito, como jogar. Para Wiener (1965, p. 173), o processo de aprendizagem presente em máquinas que jogam xadrez é possível por meio de programações específicas. Trata-se de uma programação de primeira ordem, que pode ser linear em certos casos, combinada com uma programação de segunda ordem que utiliza um segmento muito mais extenso para a determinação de como a programação de primeira ordem irá se comportar. Em outras palavras, a programação de primeira ordem seria a base lógica (contendo comandos do tipo “Se ... então”, por exemplo) para a programação de segunda ordem (que envolveria retroalimentação, por exemplo), complexificando as capacidades do sistema. Mais especificamente, a informação atual presente no sistema é utilizada para a predição do futuro por meio de uma operação não linear, mas a determinação da operação linear correta é um

problema estatístico em que o fluxo de informação passada e o passado de muitas outras informações semelhantes coletadas nas partidas anteriores são usados na base estatística. Nesse sentido, a máquina que aprende opera, necessariamente, por *feedbacks* não lineares, ou seja, operações cujos resultados podem influenciar os estados iniciais em outros momentos.

Desse modo, podemos argumentar que o projeto de máquina apresentado por Wiener mostra algumas propriedades características de sistemas complexos, tal como elucidamos nas sessões anteriores. A máquina de jogar xadrez de Wiener, a partir das informações adquiridas nas partidas, produz uma representação dos movimentos das peças e de possibilidades de jogadas. Dessa maneira, ela aprende a como jogar de uma maneira satisfatória a fim de atingir o objetivo do jogo: o xeque mate. A maneira como Wiener parece utilizar o termo *aprendizagem* se refere ao fato de que a máquina possui como programação as regras do jogo e as inúmeras combinações de jogadas, mas existe a possibilidade da máquina, a partir das abstrações dos jogos ocorridos, adquirir novas possibilidades de ação. Trata-se de uma combinação entre as regras pré-estabelecidas e as informações decorrentes da abstração da experiência de jogar. Tal combinação propicia, segundo Wiener, a emergência de novos comportamentos o que pode ser caracterizado como *aprendizagem*. A nosso ver, o projeto de Wiener abarca muitas propriedades importantes para ser caracterizado como um sistema complexo, mas não poderia ser considerado no sentido forte do termo devido ao ambiente controlado (as regras do xadrez) em que a máquina é concebida.

As máquinas de jogar xadrez lidam com um número limitado de variáveis em um ambiente controlado. Esse contexto empobrece a possibilidade do sistema tornar-se complexo no sentido forte, pois não está inserido em um ambiente não controlado, natural. Como apontamos anteriormente, as relações com o ambiente são importantes para que a informação manipulada pelo sistema se torne *significativa*, no sentido de poder alterar o comportamento e as ações do sistema em decorrência da própria dinâmica do meio. Embora Wiener afirme que sua máquina é capaz de aprender, conforme mostramos, podemos apontar que, na verdade, a máquina apenas manipula um grande fluxo de informações que podem ser equiparadas a meros símbolos destituídos de significado. Poderia, então, esse tipo de máquina ser considerada complexa? Por um lado, podemos apontar que sim, pois a máquina de jogar xadrez tal como caracteriza Wiener apresenta condições de adaptação, regulação e plasticidade de comportamentos. Todavia, por outro lado, a máquina é destituída de um critério importante para ser considerada complexa: ela parece não ser capaz de lidar com informação significativa. Sendo assim, seus processos de supostas adaptações, regulação e plasticidade se tornam apenas características aparentes e não efetivas. Argumentamos, assim,

que a capacidade de lidar com o tipo de informação significativa que norteia as relações é basilar no processo de complexificação de sistemas, sejam naturais ou artificiais. A nova questão que se coloca aqui passa a ser: sistemas artificiais podem lidar com informação significativa? Entendendo aqui a informação significativa no sentido, amplo, de envolver aspectos semânticos e pragmáticos da comunicação, além dos eminentemente sintáticos, que norteiam a ação dos agentes. Não cabe aqui aprofundar o difícil conceito de informação significativa e do sistema artificial capaz de efetivamente manipulá-la com sucesso. Nosso objetivo foi analisar projetos da IA tradicional que pretendem dar conta da concepção de máquina atuante em contextos controlados a tarefas específicas apresentada no trabalho de Weiner (1948/1965).

A proposta de Brooks, por outro lado, teoricamente consiste em construir modelos computacionais autônomos, ou seja, modelos capazes de selecionar o comportamento mais adequado a partir de sua troca de informações com o meio em que está situado o agente artificial. Isso se torna teoricamente possível graças à forma como são construídos tais agentes em termos de engenharia (arquitetura de subsunção) e de programação (algoritmos amplos o suficiente para darem margem a comportamentos não programados).

Nesse contexto, as máquinas de Brooks lidariam com informação significativa, no sentido amplo apresentado acima, na medida em que, a partir de trocas de informações com o meio, ajustam seus comportamentos. Mas será que as máquinas de Brooks, uma vez que lidam com informação significativa, podem ser conseqüentemente consideradas sistemas artificiais complexos? Teoricamente sim, pois os modelos de Brooks seriam capazes de estabelecer relações dinâmicas com o meio (pois estão inseridas nele), as quais são mediadas por relações de comunicação (troca de *feedbacks* constantes entre a máquina e o meio). Desse modo, os modelos de Brooks seriam capazes de se adaptarem e se autorregularem, apresentando, assim, um progressivo aperfeiçoamento de seus comportamentos. A problemática que colocamos aqui é referente à instanciação dos pressupostos teóricos de Brooks, pois os modelos efetivamente construídos segundo estes pressupostos não apresentaram *performances* tal como previstas na teoria.

Em suma, a conclusão a que chegamos neste trabalho é que um sistema capaz de estabelecer relações significativas com o meio (de modo a complexificar seus comportamentos, chegando a ser autônomo como pretende Brooks (1991)) deveria ser situado, incorporado, auto-organizado, complexo, dinâmico e habilidoso. Características que podem colocar em xeque o argumento funcionalista da múltipla realizabilidade proposto por Putnam (1967/1975), segundo o qual as mesmas funções podem ser instanciadas em múltiplas

estruturas físicas. A partir da limitação prática dos modelos computacionais, questionamos se, efetivamente, características como a complexidade e a habilidade podem ser satisfatoriamente instanciadas em estruturas artificiais com limitada capacidade de gerar propriedades emergentes. Os modelos como os propostos por Brooks (1991) e Beer (2003) podem ser incorporados, pois possuem um “corpo” que tenta estabelecer relações com o meio e eles são situados justamente por estarem nesse meio. Nesse contexto, eles podem também ser considerados complexos, no sentido quantitativo, pois manipulam grandes quantidades de informações. Todavia a natureza desse *corpo* parece limitar o desenvolvimento do modelo de modo que ele realmente se complexifique e manipule informação significativa a fim de ser autônomo e capaz de instanciar habilidades. Parece que um corpo não orgânico não permite que o modelo *dirigido* a informação manipulada de um modo satisfatório e que resulte na autonomia do sistema em um sentido mais abrangente. Pode ser apenas uma limitação técnica momentânea, mas consideramos que a estrutura física corpórea do sistema artificial, na qual certas características são instanciadas é altamente relevante para que o sistema possa vir a ser considerado autônomo. Assim sendo, as críticas de Dreyfus (1975) e Searle (1987) ao projeto da Inteligência Artificial ainda parecem ecoar em relação ao projeto supostamente não representacionista de vertentes incorporadas da cognição.

## CONSIDERAÇÕES FINAIS

O objetivo deste trabalho foi analisar e problematizar teses internalistas/representacionistas presentes no projeto de modelagem mecânica da mente proposta pela Ciência Cognitiva. Para tal, no Capítulo 1 focalizamos um problema clássico da filosofia, que é o de definir o conceito de *inteligência*: O que faz com que um sistema possa ser chamado de *inteligente*? Quais são os critérios que devem ser satisfeitos para que um sistema artificial possa ser considerado capaz de instanciar habilidades cognitivas? Estas interrogações ainda se fazem presentes e persistem nas problematizações e pesquisas filosóficas e cognitivistas. Muitas foram as tentativas de responder tais interrogações, a IA, por exemplo, enquanto um ramo da Ciência Cognitiva, propõe a análise da inteligência em termos de modelagem computacional. Nesse contexto, apontamos que o paradigma adotado pela IA tradicional não caracteriza satisfatoriamente a inteligência, embora proponha a caracterização mais viável tecnicamente para a modelagem (em termos de linguagem de programação e engenharia).

No Capítulo 2, procuramos mostrar que o paradigma funcional adotado pela IA tradicional, na medida em que desconsidera componentes contextuais e relacionais nos modelos cognitivos, não oferece bases para uma modelagem robótica capaz de ser contextualmente situada e de instanciar disposições que permitam adquirir e aperfeiçoar habilidades cognitivas no sentido proposto por Ryle (1949). Nesse sentido, a satisfatoriedade das habilidades está diretamente relacionada ao contexto e ao tipo de relação estabelecida entre agente e meio. Argumentamos que, para um agente ser capaz de adquirir e desenvolver habilidades, ele deve ser complexo e auto-organizado. Em outras palavras o processo de aquisição e desenvolvimento de habilidades relevantes não possui um controlador central e heterogêneo: o processo acontece espontaneamente e por si a partir das condições iniciais do meio e do sistema, que se complexifica progressivamente como decorrência da incorporação de novidades e dos processos de aprendizagem que ocorrem em suas interações com o ambiente em constante transformação. Nos perguntamos se seria possível um modelo computacional ser capaz de instanciar habilidades deste tipo e com o grau de complexidade apresentado por agentes naturais.

Nesse contexto, no Capítulo 3 investigamos se a proposta de Brooks (1990, 1991a, 1991b, 2002) de uma robótica situada e incorporada associada à abordagem ryleana de *saber como* (RYLE, 1949) e pautada na Teoria da Auto-Organização (DEBRUN, 2009) propiciaria uma modelagem que instancie propriedades disposicionais que permitam a emergência de

habilidades. E por que pautar a análise dos comportamentos inteligentes em modelos computacionais? Da perspectiva da IA fraca, podemos dizer que os modelos podem ser considerados instâncias de experimentos de pensamento e constituir a expressão efetiva de uma teoria, podendo revelar as limitações de nossa compreensão. Na perspectiva da IA forte, por outro lado, o modelo computacional seria o próprio processo cognitivo.

Mas será possível que existam máquinas capazes de complexificar-se e instanciar habilidades? A partir de conceitos da sistêmica, sugerimos que a complexidade envolve adaptação, regulação e plasticidade do sistema. Sugerimos que tais propriedades emergem de processos de comunicação no sistema que envolvem informação significativa. Ou seja, de relações dinâmicas estabelecidas entre as partes do sistema, o sistema e o meio. Por meio da comunicação, *feedbacks* ocorrem e a informação é geradora de novos processos. Tais processos permitem que o sistema evolua, ou seja, complexifique suas ações e comportamentos visando uma performance satisfatória.

Problematizamos se sistemas artificiais poderiam estabelecer relações significativas de comunicação envolvendo adaptação, regulação e plasticidade que permitissem considerá-los habilidosos no sentido proposto por Ryle (1949). Para Ryle, agentes podem adquirir e aperfeiçoar disposições habilidosas de interagir com o mundo por meio de aprendizagem e ajuste progressivo de padrões de ação. Apontamos que, em tese, é possível conceber sistemas artificiais complexos, capazes de instanciar habilidades, pois, quando assentamos a complexidade nos pressupostos anteriormente citados, não a condicionamos a estruturas físicas determinadas (orgânicas ou não). Mas, embora, em tese, seja possível a sistemas artificiais instanciar disposições e aprender habilidades, o projeto formulado por Brooks ainda parece estar longe de ser capaz de “incorporar” o tipo de habilidade complexa que os organismos aprendem cotidianamente.

Nesse sentido, argumentamos que para um sistema ser habilidoso ele deveria ser complexo, situado, incorporado e auto-organizado. Acreditamos que o *corpo* desempenhe um papel central no desenvolvimento dessas características, pois é por meio dele que a relação entre agente e meio é possível. Como dito anteriormente, em tese, é possível concebemos sistemas artificiais que instanciem habilidades, mas na perspectiva de sua viabilidade parece que a natureza do corpo situado é relevante. A artificialidade parece impor barreiras técnicas à modelagem de certas habilidades cognitivamente carregadas e que desencadeiam comportamentos inteligentes. A partir da análise dos modelos computacionais desenvolvidos até hoje, podemos questionar se é o tipo de organização da matéria orgânica que possibilita

aos sistemas serem complexos e possuírem as propriedades disposicionais capazes de propiciar o desenvolvimento de habilidades benéficas para a preservação do sistema.

Apontamos que projetos como os de Brooks e Beer podem ser considerados relevantes enquanto ferramentas explicativas, ou seja, enquanto modelos que explicam e apontam os limites das teorias e hipóteses vigentes de como os comportamentos acontecem e se eles revelam algo que pode ser considerado *inteligente*. Por fim, o próprio conceito de *inteligência* é revisto e parece descentralizar-se da concepção tradicional. Contemporaneamente, um modelo para ser considerado inteligente, deve antes de mais nada, ser adaptável. As várias profecias de desenvolvimento de sistemas artificiais inteligentes feitas desde meados do século XX de que em poucas décadas haveria tais sistemas ainda não se concretizaram, como ocorre, em geral, com todas as profecias...

Por fim, o presente trabalho ao analisar os processos cognitivos inteligentes a partir de modelos computacionais pode desencadear, como já apontamos, em uma reflexão que diz respeito à esfera do processamento lógico dos modelos e, por outro lado, à estrutura física dos mesmos. Contemporaneamente, existem muitos projetos de lógicas não clássicas (lógicas modais, polivalentes, adaptativas, algoritmos genéticos) que pretendem interpretar aspectos que não foram abordados pelas lógicas clássicas. Será que a *inteligência*, considerada enquanto *processo* de agentes situados, incorporados e auto-organizados poderia ser modelada a partir de alguma lógica não-clássica? Como ela seria instanciada? Qual o tipo de substrato físico é apto para este tipo de modelagem?

Uma sugestão de trabalho futuro seria além da análise da lógica mais apta para a modelagem de habilidades inteligentes, a análise sobre o processo de transição do físico para o semântico, ou seja, da passagem da pura manipulação sintática (as regras lógicas instanciadas em uma arquitetura física) para a emergência de um resultado que não se reduz à simples soma das partes envolvidas (o funcionamento das regras lógicas). Como se daria esse processo? E será que ele é realmente possível?

## REFERÊNCIAS

- ASIMOV, I. *Fantastic voyage II: destination brain*. New York: Doubleday, 1987.
- BABBAGE, H. P. *The Analytical Engine*. 1888. Disponível em: <<http://www.fourmilab.ch/babbage/hpb.html>>. Acesso em: 20 mar. 2012.
- BAUER, W.A. Four theories of pure dispositions. In: BIRD, A.; ELLIS, B. (Org.). *Properties, power and structures issues in the metaphysics of realism*. New York: Routledge, 2012.
- BEER, R. D. Toward the evolution of dynamical neural networks for minimally cognitive behavior. In: MAES, P. *et al.* (Ed.). *From animals to animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive*. Cambridge: The MIT Press, 1996.
- \_\_\_\_\_. The dynamics of active categorical perception in an evolved model agente. *International Society for Adaptive Behavior*, v. 11, n. 4, p. 209-243, 2003.
- BIZARRO, S. L. F. M. *Internalismo e Externalismo: um debate em filosofia da mente e da psicologia*. 1999. 115 f. Dissertação (Mestrado) - Filosofia da Linguagem e da Consciência, Faculdade de Letras da Universidade de Lisboa, Lisboa, 1999.
- BOOLOS, G. S.; BURGESS, J. P.; JEFFREY, R. C. *Computabilidade e lógica*. São Paulo: Editora Unesp, 2012.
- BRESCIANI, E; D'OTTAVIANO, I. M. L. Conceitos básicos de sistêmica. In: D'OTTAVIANO, I. M. L.; GONZALEZ, M. E. Q. (Org.). *Auto-organização: estudos interdisciplinares*. Campinas: Editora Unicamp, 2000.
- BROOKS, R. Elephants Don't Play Chess. *Robotic and Autonomous System*, v. 6, p. 3-15, 1990.
- \_\_\_\_\_. Intelligence without representation. *Artificial Intelligence*, v. 47, p 139-159, 1991a.
- \_\_\_\_\_. New Approaches to Robotics. *Science* 13, v. 253, n. 5025, p. 1227-1232, 1991b.
- \_\_\_\_\_. *Flesh and Machines: how robots will change us*. New York: Pantheon Books, 2002.
- BROOKS, R. *et al.*. The Cog Project: Building a Humanoid Robot. *Computation for Metaphors, Analogy and Agents*, v. 1562, 1998.
- \_\_\_\_\_. Humanoid robots: A new kind of tool. In: *Intelligent Systems and their Applications*, IEEE. v. 15. p. 25 – 31, 2000.
- BURGE, T. Individualism and the mental. *Midwest Studies in Philosophy*, v. 4, p. 73-121, 1979.
- CARNIELLI, W; EPSTEIN, R. L. *Computabilidade, funções computáveis e os fundamentos da Matemática*. São Paulo: Editora Unesp, 2006.

CHEMERO, A. Toward a situated, embodied realism. *Center for the Ecological Study of Perception and Action*, p. 177-204, 2007.

\_\_\_\_\_. Modeling self-organization with nonwellfounded set theory. *Ecological Psychology*, 2012.

\_\_\_\_\_. *Radical Embodied Cognitive Science*. Cambridge: MIT Press, 2009.

CLARK, A. *Being There: Putting Brain, Body and World Together Again*. Cambridge: MIT Press, Bradford Books, 1997.

\_\_\_\_\_. An embodied cognitive science? *Trends in Cognitive Sciences*, v. 3, p. 345-351, 1999.

\_\_\_\_\_. Artificial Intelligence and the Many Faces of Reason. In: STICH, S.; WARFIELD, T. (Ed.). *The blackwell guide to philosophy of mind*. Blackwell, 2003.

\_\_\_\_\_. *Supersizing the Mind: embodiment, action and cognitive extension*. *Oxford Scholarship*, 2008.

DEBRUN, M. A ideia de auto-organização. In: D'OTTAVIANO; GONZALEZ, M. E. Q. (Org.). *Michel Debrun: Identidade Nacional Brasileira e Auto-organização*. Campinas: UNICAMP; CLE, 2009.

DESCARTES, R. *Discurso do Método*. São Paulo: Nova Cultural, 1987.

\_\_\_\_\_. *Meditações metafísicas*. São Paulo: Nova Cultural, 1988.

DRYFUS, H. *O que os computadores não podem fazer: uma crítica da Razão Artificial*. Rio de Janeiro: A Casa do Livro Eldorado, 1975.

\_\_\_\_\_. *What computers still can't to do*. New York: Haper&Row, 1993.

DU, S., TAO, Y.; MARTINEZ, A. M. Compound facial expressions of emotion. *Proceedings of the National Academy of Science of the United States of America*. 2014. Disponível em: <<http://www.pnas.org/search?fulltext=Aleix+Mart%C3%ADnez&submit=yes&x=0&y=0>>. Acesso em: 1 abr. 2014.

DUPUY, J. P. *Nas origens das Ciências Cognitivas*. Unesp: São Paulo, 1996.

FETZER, J. H. *Filosofia e Ciência Cognitiva*. Bauru: Edusc, 2000.

FRANKLIN, S.; GRAESSER, A. Is it an Agent, or Just a Program?: a Taxonomy for Autonomous Agents. In: THE WORKSHOP ON INTELLIGENT AGENTS III, AGENT THEORIES, ARCHITECTURES, AND LANGUAGES, 1996, Budapest. *Proceedings...* Berlin: Springer, 1997.

GONZALEZ, M.E.Q.; NASCIMENTO, T.C.A.; HASELAGER, W.F.G. Informação e conhecimento: notas para uma taxonomia da informação. In: GONZALEZ, M.E.Q.; FERREIRA, A.; COELHO, J. (Org.). *Encontro com as Ciências Cognitivas IV*. São Paulo: Cultura Acadêmica, 2004. v. IV, p. 195-220.

GONZALEZ, J. A. Q.; BROENS, M. C.; GONZALEZ, M. E. Q.; KOBAYASHI, G. Autonomous robots and human social actions: A view from the complex systems perspective. In: *Proceedings of the 14th. European Conference on Complex Systems*. Springer, no prelo.

HASELAGER, W. F. G. Auto-organização e Comportamento Comum opções e problemas? In: SOUZA, G. M.; D'OTTAVIANO, I. M. L.; GONZALEZ, M. E. Q. (Org.). *Auto-organização: estudos interdisciplinares*. Campinas: CLE, v. 38, p. 213-135, 2004.

\_\_\_\_\_. O Mal Estar do Representacionismo: as sete dores de cabeça da Ciência Cognitiva. In: FERREIRA, A.; GONZALEZ, M. E. Q.; COELHO, J. G. (Org.). *Encontro com as Ciências Cognitivas*. São Paulo: Cultura Acadêmica, 2004. v. 4.

KHALFA, J. *A natureza da inteligência*. São Paulo: Editora Unesp, 1996.

KIRSH, D.; MAGLIO, P. *On distinguishing epistemic from pragmatic action*. 1994. Disponível em: <<http://adrenaline.ucsd.edu/kirsh/publications.html>>. Acesso em: 23 nov. 2013.

LAU, J.; DEUTSCH, M. *Externalism About Mental Content*. 2010. Disponível em: <<http://plato.stanford.edu/archives/fall2010/entries/content-externalism/>>. Acesso em: 13 out. 2012.

LECLERC, A. A concepção externalista de pessoa. In: BROENS, M. C.; MILIDONI, C. B. (Org.). *Sujeito e Identidade Pessoal: estudos de Filosofia da Mente*. São Paulo: Cultura Acadêmica, 2003.

MALONEY, J. C. Functionalism. *The MIT Encyclopedia of the cognitive sciences*. Cambridge: The MIT Press, 1999.

MINSKY, M. *The society of mind*. New York: Simon and Schuster; 1986

\_\_\_\_\_. Logical versus analogical or symbolic versus connectionist or neat versus scruffy. *AI Magazine*, v. 12, n. 2, p. 34-51, 1991. Disponível em: <<https://www.aaai.org/ojs/index.php/aimagazine/article/view/894/812>>. Acesso em: 10 jan. 2014.

\_\_\_\_\_. (1995). Matter, mind and models. *Proc. International Federation of Information Processing Congress*, v. 1, p. 45-49, 1965. Disponível em: <<http://groups.csail.mit.edu/medg/people/doyle/gallery/minsky/mmm.html>>. Acesso em: 2 dez. 2013.

MORAES, J.A. *Implicações éticas da “virada informacional na filosofia”*. Uberlândia: EDUFU, 2014.

MURPHY, R.R. *Introduction to AI Robotics*. Cambridge: The MIT Press, 2000.

NEWELL, A.; SHAW, J. C.; SIMON, H. A.; Empirical explorations with the logic theory machine: A case study in heuristics. In: FEIGENBAUM E.A.; FELDMAN, J. (Eds.), *Computers and thought*. New York: McGraw-Hill, 1963.

PUTNAM, H. (1967) The nature of mental states. In: PUTNAM, H. *Mind, language and reality*. Cambridge: Cambridge University Press, 1975.

- \_\_\_\_\_. *Representation and Reality*. Cambridge: MIT Press, 1987.
- \_\_\_\_\_. The Meaning of a Meaning. In: GEIRSSON H. ; LOSONSKY, M. (Org.). *Readings in language & mind*. 1975.
- RYLE, G. *The concept of Mind*. Londres: Hutchinson, 1949.
- SEARLE, J. R. *A redescoberta da mente*. São Paulo: Martins Fontes, 1997.
- \_\_\_\_\_. *Mente Cérebro e Ciência*. Lisboa: Edições 70, 1987.
- \_\_\_\_\_. Minds, Brains, and Programs. *The Behavioral and Brain Sciences*, v. 3, 1980.
- SHANNON, C. A Mathematical Theory of Communication. *The Bell System Technical*, v. 27, p. 379–423, 1948.
- SHAPIRO, L. The embodied research program. *Philosophy Compass*, v. 2, p. 338-346, 2007. Disponível em: <[http://portal.uni-freiburg.de/cognition/lehre/archiv/ss13/mat-prosem/embodiedcognition/Shapiro\\_EmbodiedCognition.pdf](http://portal.uni-freiburg.de/cognition/lehre/archiv/ss13/mat-prosem/embodiedcognition/Shapiro_EmbodiedCognition.pdf)>. Acesso em: 24 fev. 2014.
- \_\_\_\_\_(Ed.). *The Routledge Handbook of Embodied Cognition*. New York: Routledge , 2014.
- STERNBERG, R. The concept of intelligence. In: *Handbook of intelligence*. Cambridge: University Press, 2000.
- \_\_\_\_\_. *What's is Intelligence?* Norwood: Ablex, 1986.
- STERNBERG, R.; LAUTREY, J; LUBART, T. Where are we in the field of intelligence, how did we get here, and where are we going? *Models of intelligence: international perspective*. Washington: American Psychological Association, 2008. p. 3-25.
- THAGARD, P. *Mente: introdução à ciência cognitiva*. Porto Alegre: ARTMED, 1998.
- THE ROYAL ACADEMY OF ENGINEERING. *Autonomous system: social, legal and ethica issues*. London: The Royal Academy of Engineering. 2009.
- TEIXEIRA, J. *Filosofia e Ciência Cognitiva*. Petrópolis: Vozes, 2004.
- \_\_\_\_\_. *Mentes e Maquinas: uma introdução à Ciência Cognitiva*. Porto Alegre: Editoras Artes Medicas, 1998.
- TURING, A. Computing machinery and intelligence. *Mind*, Oxford, v. 59, 1950.
- \_\_\_\_\_. On computable numbers, with an application to the Entscheidungs problem. *Proc. London Maths. Soc.*, v. 2, n. 42, p. 230–265, 1936.
- WEAVER, W. Science and complexity. *Classical Papers – Science and complexity E:CO*, v. 6, n. 3, p. 65-74, 2004.
- WIENER, N. (1948) *Cybernetics*. 2. ed. Cambridge: MIT Press, 1965.
- \_\_\_\_\_. O Homem e a Maquina. In: KUHNER, M. H. *O conceito de informação na ciência contemporânea: Colóquios filosóficos internacionais de Royaumont*. Rio de Janeiro: Paz e Terra, 1970. p. 69-99.

\_\_\_\_\_. *The human use of human beings: cybernetics and society*. London: Free Association Books, 1989.