

**UNIVERSIDADE FEDERAL DE SÃO JOÃO DEL REI - UFSJ
DEPARTAMENTO DE FILOSOFIA E MÉTODOS - DFIME**

GEOVANNI DE MORAIS GAVA

**AS CONTRIBUIÇÕES DA INTELIGÊNCIA ARTIFICIAL
PARA FILOSOFIA DA MENTE CONTEMPORÂNEA**

**SÃO JOÃO DEL REI - MG
2017**

GEOVANNI DE MORAIS GAVA

**AS CONTRIBUIÇÕES DA INTELIGÊNCIA ARTIFICIAL PARA
FILOSOFIA DA MENTE CONTEMPORÂNEA**

Monografia apresentada ao departamento de filosofia e métodos – DFIME/UFSJ como pré-requisito para obtenção de título de bacharel em filosofia.

Orientador: Gustavo Leal Toledo

**SÃO JOÃO DEL REI - MG
2017**

GEOVANNI DE MORAIS GAVA

**AS CONTRIBUIÇÕES DA INTELIGÊNCIA ARTIFICIAL PARA
FILOSOFIA DA MENTE CONTEMPORÂNEA**

Monografia apresentada ao departamento de filosofia e métodos – DFIME/UFSJ como pré-requisito para obtenção de título de bacharel em filosofia.

São João Del Rei, 12 de abril de 2017.

Banca Examinadora

Prof(a). Nome da Instituição de Origem

Prof(a). Nome da Instituição de Origem

Prof(a) Orientador. Nome da Instituição de Origem

AGRADECIMENTOS

À Universidade Federal de São João Del Rei.

*Podemos avistar só um pequeno trecho do
caminho à nossa frente, mas ali já vemos muito do
que precisa ser feito (Alan Turing)*

RESUMO

A filosofia da mente é uma disciplina recente em comparação aos demais conteúdos filosóficos, a partir de Descartes é possível perceber algumas tentativas de delimitação da mente no sentido de compreender qual a verdadeira natureza do pensamento. Com o desenvolvimento científico, especialmente com o advento do computador digital e da inteligência artificial, foi possível a elaboração de uma nova perspectiva para o estudo da mente sob a ótica de modelos computacionais. Essa nova perspectiva que busca uma explicação para a natureza da mente tendo como base os computadores digitais ganhou o nome de funcionalismo e é considerada hoje uma das principais teorias da mente existentes, o objetivo deste trabalho é compreender como a inteligência artificial contribuiu para o desenvolvimento da teoria funcionalista da mente.

Palavras-chave: Filosofia da Mente, Inteligência Artificial, Funcionalismo

SUMÁRIO

Introdução	7
Capítulo 1 - O problema mente e corpo: panorama das principais teorias da mente	8
1.1 <i>Dualismo de Substâncias</i>	8
1.2 <i>Dualismo de Propriedades</i>	10
1.3 <i>O Fisicalismo</i>	12
1.3.1 <i>Behaviorismo Filosófico</i>	13
1.3.2 <i>Fisicalismo Redutivo (teoria da identidade)</i>	14
1.3.3 <i>Funcionalismo</i>	16
Capítulo 2 – Contribuições da Inteligência Artificial para a Filosofia da Mente	20
2.1 <i>A Tese da Multipla Instaciação</i>	20
2.2 <i>A relação entre o Funcionalismo e a Inteligência Artificial</i>	21
2.3 <i>A Quebra do paradigma Representacionalista na Inteligencia Artificial</i>	25
2.4 <i>A Nova Abordagem</i>	29
Capítulo 3 – Considerações Finais	34
Conclusão	36
Referências Bibliográficas	37

Introdução

O presente trabalho pretende apresentar uma análise da teoria funcionalista da mente no sentido de apontar as contribuições da inteligência artificial para o desenvolvimento e consolidação desta teoria. Para os pesquisadores da inteligência artificial o funcionamento da mente humana seria análogo ao funcionamento dos computadores, de forma que seria possível replicar as capacidades da mente humana em um computador. É essa perspectiva que orienta a inteligência artificial, e que ao mesmo tempo causa um grande impacto em outras disciplinas como a filosofia da mente. A filosofia da mente tem como questão principal a busca por argumentos que demonstrem qual a real natureza da mente e qual a relação existente entre a mente e o cérebro, assim, com o advento da inteligência artificial surge também uma nova forma para o estudo da mente através de modelos computacionais, a qual foi denominada teoria funcionalista da mente. Iniciaremos no cap. 1 promovendo uma exposição geral a respeito do conteúdo abordado em filosofia da mente, traçando um breve panorama do seu desenvolvimento a partir de platão. Em seguida, no cap 2, apresentaremos as contribuições da inteligência artificial para a teoria funcionalista da mente, na tentativa de compreender como a inteligência artificial foi importante para o desenvolvimento dessa teoria, tendo como base a análise do artigo “computação e inteligência” de Alan Turing e de experimentos de

Inteligência Artificial realizados pelos pesquisadores Allen Newell e Herbert Simon, em seguida nas secções 2.3 e 2.4 apresentaremos a quebra do paradigma representacionista na inteligência artificial e como a nova abordagem pode colaborar para a consolidação e aprimoramento da teoria funcionalista da mente. Por fim, com base nos textos de Daniel Dennett, apontaremos para uma relação necessária, e benéfica para ambas disciplinas, entre a Inteligência Artificial e o Funcionalismo.

CAPÍTULO 1 - O PROBLEMA MENTE E CORPO: PANORAMA DAS PRINCIPAIS TEORIAS DA MENTE

1.1 – Dualismo de substâncias¹

Apesar de não conseguirmos saber ao certo quem foi o primeiro pensador a trabalhar o problema mente e corpo. Verifica-se que Platão foi um dos primeiros a abordar esta questão, propondo uma teoria dualista, na qual existe uma separação ontológica entre corpo e alma, sendo que, esta última existe independentemente do corpo. Em outras palavras, para Platão, a alma é indissolúvel e imortal, ou seja, eterna, ao passo que o corpo é dissolúvel e mutável. Para Platão a realidade física que encontramos no mundo é apenas uma cópia imperfeita do mundo das Formas, o qual acessamos através de nosso intelecto.

Essa concepção dualista, na qual o corpo e a alma se tratam de substâncias distintas, não é defendida apenas por Platão, assim como ele, o filósofo René Descartes também acreditava na existência de duas substâncias totalmente diferentes, a saber, a *res extensa* (substância física) e *res cogitans*²(substância mental), onde o principal atributo da substância física é sua extensão, enquanto o principal atributo da substância mental é o pensamento.

Compreendi então que eu era uma substância cuja essência ou natureza consiste somente no pensar e que, para ser, não necessita de lugar algum, nem depende de qualquer coisa material. Desse modo, esse eu, isto é, a alma, pela qual sou o que sou, é inteiramente distinta do corpo e até mesmo que ela é mais fácil de conhecer do que ele e, ainda que esse nada fosse, ela não deixaria de ser tudo o que é. (DESCARTES, 2009, p.42)

O dualismo cartesiano trata-se de um dualismo interacionista, pois essas duas substâncias, apesar de serem distintas, estão em constante interação.

¹ Dualismo ontológico ou dualismo de realidades.

² O conceito de *res extensa* cartesiano está relacionado a toda e qualquer coisa física existente no mundo, inclusive seu próprio corpo. O conceito de *res cogitans* cartesiano está relacionado substância pensante, ou a mente. Sendo esta uma substância totalmente incorpórea e imaterial.

A natureza me ensina, também, por esses sentimentos de dor, fome, sede etc., que não somente estou alojado em meu corpo, como um piloto em seu navio, mas que, além disso, lhe estou conjugado muito estreitamente e de tal modo confundido e misturado, que componho com ele um único todo (DESCARTES, 1973, p. 144).

É neste ponto, onde ocorre a interação entre as substâncias, que se insere a crítica ao dualismo cartesiano. Descartes já tinha conhecimento a respeito da lei de conservação da quantidade do movimento, sendo assim, “a matéria deve se comportar no espaço seguindo leis rígidas, e não se pode obter movimento corporal (= quantidade de movimento) a partir do nada” (CHUCRLAND, 2004, P.28), em outras palavras, sendo o pensamento uma substância imaterial, como esta pode ter efeito causal sobre a substância física? Descartes, na tentativa de defender seu dualismo interacionista, recorre à *glândula pineal* e aos *espíritos animais* para explicar como essa interação acontece. Para ele os espíritos animais são os responsáveis pela comunicação entre alma e corpo, e a glândula pineal é onde a alma reside no corpo. Porém esses argumentos não foram de todo convincentes, e mesmo em sua época já existiam críticas a respeito de como ocorria de fato essa interação.

Na sua obra *O Discurso do Método* de 1637, em sua parte Quinta, Descartes fala acerca do movimento do coração e das artérias, para ele o funcionamento do corpo humano é parecido com o funcionamento de um relógio no qual cada engrenagem desempenha um papel específico.

Finalmente, o que há de mais notável em tudo isso [movimento do coração e das artérias] é a geração dos espíritos animais que são como um vento muito sutil, ou melhor, como uma chama muito pura e muito viva que, subindo ininterruptamente em grande quantidade do coração ao cérebro, daí se dirige pelos nervos para os músculos e confere movimento a todos os membros (DESCARTES, 2009, p. 58-59).

Dentre todos os órgãos do corpo humano ele destaca a importância do coração e artérias, que devido ao seu funcionamento produz e leva calor para cérebro (glândula pineal), fazendo com este produza os espíritos animais e os remeta novamente aos membros, através dos nervos, possibilitando assim as sensações e a locomoção.

Na teoria cartesiana, a glândula pineal está no centro da explicação dos movimentos e das sensações, que são explicados a partir do movimento dos espíritos animais e dos nervos que os transportam. (...) A origem do movimento dos espíritos e do sangue encontra-se na ação do coração, em outras palavras, o princípio corporal dos

movimentos de nossos membros consiste no fogo cardíaco que, segundo a concepção cartesiana, é mantido pelo sangue das veias. (DONATELLI, 2003, p.82)

Hoje, esse tipo de dualismo de substância não possui aceitação no campo da ciência de forma geral, e no campo da filosofia da mente de forma específica, uma vez que, se tratarmos a glândula pineal como substância física teríamos sérias dificuldades em explicar o relacionamento desta com a alma (substância pensante); por outro lado, se tratássemos ela como substância imaterial, teríamos a mesma dificuldade de explicar sua relação com o corpo (substância física). Em virtude de tais limitações que impedem o desenvolvimento científico da referida teoria, essa concepção dualista foi se modificando e se transformando com objetivo de buscar uma base empírica mais sólida para construção de uma ciência que de conta do problema mente-corpo.

Verificamos que o modelo científico atual exigiu que a mente deixasse de ser analisada enquanto substância imaterial³ e incorpórea que interage com a substância física, como pensava Descartes. Essa exigência teórica fez com que o problema mente-corpo fosse abordado de outra forma, na qual a mente é analisada como uma propriedade do cérebro, conforme veremos abaixo.

1.2 – Dualismo de propriedades

Considerando o desenvolvimento das ciências físicas e a atual concepção de mundo, na qual se acredita que todo o universo é constituído por partículas físicas, podemos afirmar que o dualismo cartesiano não é mais tão plausível quanto era a sua época. Em virtude disso, uma nova análise do problema mente-corpo foi sendo elaborada com o objetivo de localizar a mente no mundo físico. Assim a mente passa ser analisada como algum tipo de produto ou resultado do funcionamento do cérebro, apesar de possuir propriedades específicas das quais nenhuma outra substância física possui. Essa nova concepção, chamada de dualismo de propriedades engloba um conjunto grande de teorias, segundo Churchland, todas essas teorias tem uma idéia básica, a saber:

A idéia básica das teorias que discutimos sob este título é a de que, embora não haja uma outra *substância* envolvida nessa questão, além do cérebro físico, o cérebro é

³ O conceito de Imaterial é utilizado aqui como sinônimo de não localidade espacial, ou melhor, aquilo que não pode ser localizado no espaço-tempo.

dotado de um conjunto especial de *propriedades* das quais nenhum outro tipo de objeto físico dispõe. Essas propriedades especiais são não-físicas: daí o título *dualismo de propriedades*. (CHURCHLAND, 2004, p.30).

Desta forma, o dualismo ainda se faz presente, agora se apresentando em uma nova forma, na qual o cérebro possui propriedades físicas (estados cerebrais) e propriedades não-físicas (estados mentais), onde estas últimas, não podem ser reduzidas a partir do conhecimento científico atual. Uma das principais diferenças dessa nova versão dualista é que os estados mentais não existem independentemente do corpo, dessa forma, ao contrário do dualismo de substâncias, não pode haver a mente sem que haja antes um substrato físico para instanciá-la.

Churchland aponta o epifenomenalismo⁴ como à forma mais antiga de dualismo de propriedades, onde a mente será tratada como um fenômeno que está “além” dos processos físicos que ocorrem no cérebro.

o epifenomenalismo. Esse termo impressiona, mas seu significado é simples. O prefixo grego “epi” significa “acima”, e essa posição afirma que os fenômenos mentais não constituem uma parte dos fenômenos físicos no cérebro determinado, em última análise, nossas ações e comportamentos, mas, ao contrário, eles ocorrem, por assim dizer, “acima do embate” . (CHURCHLAND, 2004, p. 31).

O epifenomenalismo trata a mente como um sub-produto dos processos cerebrais, fazendo com que ela não possua poderes causais sobre o nosso comportamento ou nossos processos cerebrais, propondo que a causa dos fenômenos mentais seria processos extremamente complexos que são realizados no cérebro. Assim, os fenômenos mentais são produzidos pelo cérebro, e a experiência qualitativa oriunda de tais fenômenos é irreduzível aos processos cerebrais que a causou, além de não possuir poderes causais nenhum.

Para tornar o problema ainda mais complicado, as *qualia*⁵ são muitas vezes tidas como causalmente ineficazes, ou seja, elas não alterariam em nada nosso

⁴ É importante esclarecer acerca das diversas formas de epifenomenalismos existentes, algumas delas dualistas e outras fisicalistas, destaco que o epifenomenalismo que se refere o presente texto trata-se de uma forma dualista na qual existem duas propriedades ontologicamente diferentes, a saber, os processos cerebrais e os epifenômenos que contêm em si próprios experiências qualitativas irreduzíveis ontologicamente.

⁵ Entenda-se Qualia como a experiência qualitativa oriunda dos epifenômenos.

comportamento ou o funcionamento do nosso cérebro. À primeira vista isto pode parecer muito estranho, como nossa experiência consciente não pode alterar o nosso comportamento? O problema aqui é que as teorias em que as *qualia* alteram o comportamento chegaram a um “beco sem saída”.(LEAL-TOLEDO,2005, p.15)

Existem outras formas de epifenomenalismo, na qual os epifenômenos possuem poderes causais sobre o nosso comportamento, porém o risco em defender esta posição juntamente com o dualismo de propriedades sem que cometamos o mesmo erro encontrado na filosofia dualista interacionista de Descartes é muito grande, uma vez que, da mesma forma que o dualismo cartesiano encontra dificuldades para explicar como ocorre essa interação entre as substâncias, o dualista de propriedades também incorreria nas mesmas dificuldades ao afirmar que os epifenômenos possuem poderes causais sobre os processos físicos no cérebro. Em outras palavras, afirmar que os processos físicos que ocorrem no cérebros são causados por fenômenos não-físicos cuja principal característica é o aspecto qualitativo irreduzível, seria o mesmo que recorrer a algo como a glândula pineal ou os espíritos animais, e é este o “beco sem saída” que se pretende evitar.

1.3– O Fisicalismo⁶

O fisicalismo trás uma nova perspectiva para solução do problema mente-corpo, sua idéia central é: mente e cérebro são a mesma coisa. Para um fisicalista a realidade é composta apenas de uma única substância, a substância física. Segundo a doutrina fisicalista o problema mente-corpo deve ser analisado de forma simples, mente é igual a corpo, e não existe nada além do físico. Jaegwon Kim nos apresenta o fisicalismo da seguinte forma “tudo que existe no mundo espaço-temporal é uma coisa física, e todas as propriedades das coisas físicas são ou propriedades físicas ou propriedades intimamente relacionadas à sua natureza física.”. É importante ressaltar que existem varias correntes fisicalistas, e que o objetivo principal deste texto é traçar um panorama rápido e geral das teorias consideradas fisicalistas. No primeiro momento irei apresentar a teoria de Gilbert Ryle, o behaviorismo filosófico, que pretende reduzir estados mentais a estados comportamentais. Em seguida, apresentarei a perspectiva

⁶ O termo Fisicalismo foi utilizado como sinônimo de Materialismo, uma vez que as diferenças dos referidos conceitos não são relevantes para a compreensão das idéias contidas neste texto.

reduzida (teoria da identidade), e por fim tratarei sobre o funcionalismo, uma das principais teorias fisicalista atuais.

1.3.1– Behaviorismo Filosófico

Conforme foi exposto acima o fisicalismo descarta totalmente a possibilidade da existência de substâncias imateriais, ou não-físicas, sendo que a linguagem que deverá ser utilizada para todas as ciências deve ser a linguagem da física (leis da física), pois é a física que irá garantir o caráter científico daquela ciência em questão. Restou, portanto, para os fisicalistas, explicarem algumas coisas, por exemplo: como utilizar as leis da físicas para medir eventos mentais ? Como é possível reduzir teoricamente os eventos mentais utilizando para isso as leis da física? Na tentativa de responder estes questionamentos Gilbert Ryle, desconsiderando todas as formas de dualismo existentes até o momento, propõe um estudo da mente humana baseado apenas em padrões comportamentais, em seu livro *The concept of mind*, considerado por muitos como marco inicial da filosofia da mente contemporânea, ele apresenta sua teoria comportamental analítica que tem como objetivo reduzir o vocabulário da psicologia popular a padrões comportamentais, onde estes últimos, por sua vez, permitem uma verificação quantitativa.

Para Ryle, não existe uma mente que está acima e além do corpo, e é por isso que ele descreve de maneira irônica e pejorativa o dualismo cartesiano como sendo uma doutrina do fantasma (alma) na máquina (corpo). Para o behaviorista analítico, *todo* o vocabulário da Psicologia Popular – isto é, os termos que dizem respeito às nossas sensações, crenças e desejos – pode e deve ser reduzido a *padrões comportamentais*(LIMA, 2010, p. 23)

A principal crítica feita ao behaviorismo filosófico é que este não leva em consideração o aspecto qualitativo dos estados mentais, ou seja, um behaviorista não leva em considerações, por exemplo, a experiência de sentir dor ou enxergar a cor vermelha, uma vez que essas experiências não estão disponíveis, ou não poder ser retiradas, através da observação do comportamento.

Sendo assim, após descartar o estudo da natureza interior dos estados mentais, a teoria behaviorista filosófica tem como objeto de estudo a análise do vocabulário que usamos para falar acerca desses estados (cf. Churchland, 2004). Em outras palavras, uma análise apurada

do vocabulário que utilizamos para falar sobre nossos estados mentais daria conta de resolver o problema mente-corpo. E é por isso que o problema mente-corpo é considerado um problema de interpretação lingüística, ou pseudo problema dentro desta teoria.

Em sua forma mais radical e simples, o behaviorismo filosófico afirma que toda sentença sobre estados mentais pode ser parafraseada, sem perda de significado, numa longa e complexa sentença sobre que o comportamento observável *iria* resultar se a pessoa em questão estivesse nesta, naquela ou em outra circunstância observável.(CHURCLAND, 2004, p. 49)

Por fim, é importante lembrar o momento histórico em que Ryle escreveu sua obra, uma vez que nesse mesmo período havia fortes críticas tanto a psicologia introspeccionista, quanto a visão dualista de propriedades, isso faz com que Ryle deixe totalmente de lado qualquer tipo de estudo com relação às experiências qualitativas da mente, e desconsidere o aspecto interior dos estados mentais.

1.3.2 – Fisicalismo reduutivo (teoria da identidade)

Surge no final dos anos 50 a teoria materialista reducionista chamada teoria da identidade, onde o principal argumento apresentado era que: os estados mentais são estados físicos do cérebro, e que existe uma identidade de tipo entre estados mentais e processos físicos cerebrais. Por isso se faz necessário o avanço nas ciências que estudam o funcionamento e a fisiologia do cérebro, pois, mais tarde, o conhecimento adquirido acerca desse funcionamento nos servirá de base para estabelecer uma identidade perfeita entre fenômenos mentais e processos neurobiológicos.

Se faz necessário lembrar que os adeptos da teoria da identidade, ao contrário do behaviorismo filosófico, não afirmam que exista uma identidade semântica entre a psicologia popular e processos cerebrais.

os defensores da teoria de identidade não afirmam que há uma identidade semântica entre os termos da Psicologia Popular e da Neurobiologia. Isto é, a frase “estou com dor” (vocabulário da Psicologia Popular⁷) não tem o mesmo significado de “minhas

⁷ A psicologia popular, ou folk psychology é uma expressão cunhada por Dennett em 1981 e designa uma teoria habitual que todos nós possuímos por meio da qual explicamos os comportamentos de outros seres humanos recorrendo as idéias comuns de “intenção”, “crença”, “desejo” e outros termos que compõem o chamado vocabulário mentalista. (TEIXEIRA, 2008, p.33).

fibras C estão sendo estimuladas” (vocabulário neurocientífico) (LIMA, 2010, p.24).

Porém, a impossibilidade de estabelecer uma identidade semântica, entre os termos citados acima, por exemplo, “estou com dor” e “minhas fibras C estão sendo estimuladas”, não implica necessariamente em uma diferença entre eles. Pelo contrário, os adeptos a teoria da identidade acreditam que com o avanço da neurociência poderemos afirmar que o termo “estou com dor” é idêntico a “minhas fibras C estão sendo estimuladas” da mesma forma que o termo “água” é idêntico a “H²O” . Esse tipo de redução é chamado de redução interteórica, onde se utiliza de uma nova teoria, pelo fato desta ter um poder explicativo mais abrangente, para explicar um mesmo fenômeno que já possuía uma definição (menos abrangente).

Aprendemos que a luz nada mais é que ondas eletromagnéticas, e nossa melhor explicação atual diz que a cor de um objeto é idêntica a uma trinca de coeficientes de refletância que o objeto tem, como se ele estivesse “tocando uma corda musical”, embora as “notas” sejam tocadas em ondas eletromagnéticas, e não em ondas sonoras. (CHURCHLAND, 2004, p53)

A teoria da identidade pretende estabelecer o mesmo raciocínio citado acima com relação aos estados mentais identificando-os aos seus processos cerebrais correspondentes. Porém, segundo os próprios adeptos da teoria da identidade, até o presente momento não possuímos conhecimento suficiente para tanto. Por fim, a teoria materialista reducionista da identidade apresentar fortes argumentos a favor do materialismo em relação ao dualismo, uma vez que dentro de sua fundamentação não existe lugar para qualquer tipo de experiência qualitativa ontologicamente subjetiva. Porém, por outro lado, se mostra frágil em relação aos argumentos apresentados por outras formas materialistas rivais, como o funcionalismo, conforme veremos em seguida.

1.3.3 – Funcionalismo

A teoria funcionalista traz sua contribuição para o estudo da mente em meio a um cenário de pesquisas relacionadas à inteligência artificial nas décadas de 60 e 70, tendo como marco inicial o artigo publicado em 1950 pelo pai da ciência da computação, Alan Turing, com o título “Computing Machinery and Intelligence”, traduzido para o português com o título de “Computação e Inteligência” nesse artigo Turing propõe a seguinte questão: podem

os computadores pensar ? Para responder essa questão ele propõe um teste, no qual computadores serão submetidos a jogo de perguntas, chamado jogo da imitação⁸. O objetivo principal deste jogo é descobrir se um computador pode se passar por um ser humano, na medida em que ele emite respostas válidas as perguntas realizadas. Caso ele consiga responder as perguntas de forma idêntica a um humano, ele terá passado no teste e conseqüentemente, segundo Turing, não haveria problemas em atribuir a esse computador estados mentais.

A premissa fundamental de Turing (...) é a seguinte: se o comportamento de uma máquina for indistinguível daquele de um ser humano, não há nenhuma razão que nos impeça de atribuir a essa maquina *estados mentais*. O jogo da imitação é um caso específico do comportamento linguístico.(TEIXEIRA, 2008, p.32).

O critério utilizado pelo jogo da imitação para verificar se algo ou alguém pensa é pragmático e funcional, de forma que, se uma máquina funciona de maneira idêntica a um ser humano, não há por que não atribuir estados mentais a mesma. Diversas críticas foram lançadas a este ao teste de turing, e com certeza existem situações as quais Turing não levou em consideração ao elaborar o teste, o que possibilitou o desenvolvimento de várias críticas, conforme veremos na próxima seção ao tratar da crítica a Inteligência artificial forte realizada por Jonh Searle. De toda forma, pela primeira vez, após muitos tempo de estudos relacionados à mente e ao cérebro, graças a Turing, foi possível o desenvolvimento do computador digital e conseqüentemente da inteligência artificial, os quais de uma forma ou de outra, contribuíram e continuam contribuindo, com os estudos da mente, como também de outras áreas⁹ do conhecimento.

Após essa breve descrição a respeito do nascimento do funcionalismo, vamos abordar o seu principal argumento, a tese da *realização múltipla*. De acordo com a teoria funcionalista, diferentes substratos físicos podem produzir estados mentais, ou seja, o que importa para existência dos estados mentais não é a matéria do qual o cérebro é composto, mas sim as relações funcionais internas que ele mantém. Assim, de acordo com essa

⁸ O jogo da imitação ou teste de Turing será abordado de forma completa na próxima seção onde trataremos da crítica de John Searle a IA forte.

⁹ Outras áreas do conhecimento como a medicina, geologia, e matemática utilizam os chamados sistemas especialistas, para trabalhar com operações complexas dentro de suas áreas de conhecimento, para maiores informações acerca de sistemas especialistas, conferir MYCIN, PROSPECTOR e HEURÍSTICO em (TEIXEIRA, 1998, p. 51).

argumentação, percebemos que existem outros tipos infinitos de materiais que podem instanciar uma mente como a nossa, desde que estes tenham uma economia funcional interna idêntica a nossa. Dessa forma podemos notar que a abordagem funcionalista entra em conflito com a teoria da identidade, uma vez que não existe apenas um tipo de substrato específico que se relaciona com um determinado estado mental específico, pelo contrário, existem tantos tipos de substratos físicos capazes de causar estados mentais que seria impossível existir uma correspondência um-a-um entre tipos físicos e tipos mentais.

Por um lado, o funcionalismo coloca em xeque a identidade tipo mental = tipo físico, por outro lado, continua aceitando algum tipo de redução, ou identidade entre espécie mental = espécie físico, nas palavras de Churchland:

Embora os funcionalistas rejeitem a teoria da identidade “tipo mental = tipo físico”, virtualmente todos eles permanecem comprometidos com a teoria da identidade mais branda “espécie mental = espécie físico”.(CHURCHLAND, 2004, p. 70).

Essa passagem demonstra que, ao se admitir a abordagem funcionalista, admite-se também, mesmo que de forma tácita, algum tipo de redução, uma vez que os estados mentais não são totalmente independentes em relação ao físico, eles não possuem existência própria, pelo contrário, para que eles existam é necessário algum espécie físico para instanciá-los.

Neste ponto, gostaria de ressaltar que a teoria funcionalista é compatível tanto com a abordagem dualista de propriedades quanto com a abordagem materialista.

A perspectiva funcionalista está de acordo com um dos principais argumentos do dualismo de propriedades, que diz: mesmo possuindo todo o conhecimento acerca das propriedades e reações químicas que ocorrem no cérebro, seria impossível descrever estados mentais como “sentir dor”, pois, o que importa para compreensão desse estado mental não é o conhecimento acerca de sua composição química, mas sim o seu papel funcional que é determinado pelo “conjunto de relações causais que ele mantém com (1) os efeitos do meio ambiente sobre o corpo, (2) com outros estados mentais e (3) com o comportamento corporal” (CHURCLAND, 2004, p. 67).

Sendo assim não basta o conhecimento acerca das substâncias físicas que compõem o cérebro, ainda é necessário conhecer a economia funcional de seus estados internos, ou melhor:

Pouco adianta eu estudar a composição química específica do cérebro, ela não me dará a chave para que eu possa compreender o funcionamento mental, da mesma maneira que a composição química das peças de marfim num jogo de xadrez não me permite entender suas funções e regras do jogo. (TEIXEIRA, 2000, p.124).

Teixeira ao comparar o funcionamento de um jogo de xadrez que opera utilizando um tabuleiro e um conjunto de regras e funções das peças, com o funcionamento da mente humana, que opera utilizando o cérebro como tabuleiro e a mente como o conjunto de regras e funções internas do cérebro, nos mostra que a teoria funcionalista é compatível com a teoria dualista de propriedades. Porém, por outro lado ele nos adverte:

o funcionalismo não implica necessariamente uma postura materialista mas também não é incompatível com este último. Um aparelho de rádio (hardware) toca uma música (software): a música e o aparelho de rádio são coisas distintas, irreduzíveis uma a outra, embora sejam ambas necessárias para que possamos ouvir uma música. A música (ondas eletromagnéticas) é diferente do aparelho de rádio (hardware), mas ambos fazem parte do mundo material. Neste sentido, podemos sustentar a compatibilidade do funcionalismo com o materialismo, uma visão preferida pelos filósofos da mente que repensaram o problema mente-cérebro à luz da Inteligência Artificial. (TEIXEIRA, 1998, p.49).

Assim, os adeptos da teoria funcionalista não refutam totalmente o materialismo, eles apenas preferem outro nível de explicação que não leva em consideração o substrato físico no qual a mente é instanciada. Esse nível de explicação utilizado reforça ainda mais a analogia entre mentes e máquinas, onde a mente ou o pensamento seria o software (conjunto de instruções de máquina ou algoritmo) e o cérebro seria o hardware (substrato físico que executa as instruções fornecidas pelo software).

A analogia supracitada apresenta um problema, pois com base nos estudos da neurociência atual, sabemos que nosso cérebro funciona devido a ligações realizadas entre os neurônios, através do envio e recebimento de neurotransmissores¹⁰, possibilitando assim a transmissão das informações, isso equipa o cérebro com uma plasticidade capaz de se adaptar de acordo com as circunstâncias ambientais, em outras palavras, a estrutura do cérebro permite que essas ligações entre neurônios sejam feitas e desfeitas de diversas formas, de acordo com as circunstâncias ambientais. Entretanto, ao analisarmos a estrutura de um

¹⁰ Substância química produzida pelos neurônios, com objetivo de enviar informações para outras células.

hardware de computador verificamos justamente o contrário, não existe nenhum tipo de plasticidade no que se refere às ligações entre os chips que o compõe, existindo apenas as ligações já pré-estabelecidas. Conseqüentemente a analogia: a mente está para o corpo, assim como o software está para o hardware fica prejudicada.

O funcionalismo, apesar de possuir várias objeções, principalmente no que tange a analogia mente/software, trouxe a baila uma perspectiva metodológica totalmente diferente das demais teorias, ao levar o estudo da mente para um campo teórico/abstrato no qual os estados mentais são analisados apenas em função de seus papéis funcionais, visando a possibilidade de uma possível transposição da mente para outros tipos de materiais. Contudo sabemos das dificuldades em realizar tal proeza, e não podemos afirmar que com base na teoria funcionalista conseguiremos um dia realizar tal feito. Porém não podemos negar o sucesso cada vez maior nas pesquisas relacionadas à inteligência artificial principalmente no que tange ao desenvolvimento de sistemas especialistas, ou sistemas de inteligência artificial fraca – I. A. fraca¹¹ cuja implementação só foi possível mediante o uso da metodologia proposta pela teoria funcionalista.

¹¹De acordo com a IA no sentido fraco, o principal valor do computador para o estudo da mente reside no fato de que este nos fornece uma ferramenta extremamente poderosa. (SEARLE, 1996, p. 64).

CAPÍTULO 2 – CONTRIBUIÇÕES DA INTELIGÊNCIA ARTIFICIAL PARA A FILOSOFIA DA MENTE

2.1 - A tese da múltipla instanciação

O funcionalismo nos diz que uma mente é uma mente porque funciona como tal, pouco importa o substrato físico que venha instanciá-la. Dessa forma, caso fosse possível criar um dispositivo utilizando-se de qualquer tipo de substrato físico (seja matéria orgânica ou chips de silício) que funcionasse de maneira idêntica a uma mente, não haveria por que negar que esse dispositivo tenha de fato uma mente. Esse é o elo principal que liga o funcionalismo à inteligência artificial, uma vez que o objetivo principal da inteligência artificial é a produção de um sistema artificial que demonstre capacidades funcionais idênticas àquelas encontradas na mente humana. O funcionalismo é uma das principais vertentes que encontramos na filosofia da mente contemporânea e sua relação com a inteligência artificial (objeto de estudo desse projeto) se dá mediante a tese da múltipla instanciação que encontramos nos computadores digitais.

O aspecto mais interessante do funcionalismo é sua característica não-reducionista, do qual podemos derivar a chamada tese da múltipla instanciação (multiplerealizability). De acordo com essa tese, dois computadores podem diferir fisicamente um do outro, mas isso não impede que eles possam rodar o mesmo software. Inversamente, dois computadores podem ser idênticos do ponto de vista físico, mas realizar tarefas inteiramente distintas se o software for diferente. (TEIXEIRA, 2008, p.48).

De acordo com Teixeira as características mais interessantes do funcionalismo é seu caráter não-reducionista que se relaciona diretamente com a tese da múltipla instanciação, extraída da teoria da computação, que garante a autonomia existente entre hardware e software. O que nos mostra ser possível, a partir de uma perspectiva materialista, inspirada na própria arquitetura dos computadores, propor uma explicação para natureza da mente sem recorrer a qualquer tipo de substância ou princípio metafísico. Assim, considerando que o atributo principal da mente é função que ela desempenha, a teoria funcionalista não se compromete com nenhum tipo de redução da mente a algum substrato físico, pelo contrário, existem inúmeros substratos físicos que poderiam ser organizados de maneiras diferentes com o objetivo de produzir respostas funcionais idênticas àquelas produzidas pela mente humana. Considerando a

analogia entre mentes e máquinas, na qual a mente corresponde ao software e o cérebro corresponde ao hardware, é possível sustentar que, o que garante a existência da mente é seu papel funcional e não sua organização física interna. Tal analogia pode ser muito proveitosa para ambas as disciplinas, pois, por um lado a inteligência artificial precisa defender prática e teoricamente a reprodução da inteligência humana em máquinas, esse é um de seus principais objetivos desde seu surgimento, por outro lado o funcionalismo, como teoria filosófica da mente, busca uma explicação, em qualquer nível (teórico ou prático) desde que contenham argumentos coerentes, da verdadeira natureza da mente. Nesse ínterim, percebemos que é o computador digital a ferramenta principal utilizada pelos funcionalistas, a tese da múltipla instanciação reforça a analogia entre mentes e máquina, incitando a idéia de que um mesmo substrato físico (cérebro ou computador) pode executar funções diferentes, ao passo que funções idênticas podem ser executadas em um substrato físico diferente.

2.2 – A relação entre o Funcionalismo e a Inteligência Artificial

O critério utilizado pelo jogo da imitação para verificar se algo ou alguém pensa é pragmático e funcional, de forma que, se uma máquina funciona de maneira idêntica a um ser humano, não há por que não atribuir estados mentais a mesma. Turing é o primeiro a estabelecer um critério funcional como delimitador da mente, ou do pensamento, isso ocorre por que ele percebe que seria impossível, pelo menos por hora, uma definição teórica sobre aquilo que realmente constitui o pensamento. Os trabalhos desenvolvidos por Turing possibilitaram uma interseção entre o funcionalismo a inteligência artificial, e através dos computadores surge uma nova forma para o estudo da mente. O pressuposto levantado por Turing de que a mente pode ser instanciada tanto por um cérebro como por uma máquina, levando em consideração o critério funcional, abriu as portas para o grande desenvolvimento da inteligência artificial, motivando a elaboração de diversos experimentos na área. Turing pode ser considerado um dos primeiros pensadores funcionalistas, apesar de Turing não contar com uma teoria da mente, o critério funcional sugerido por ele veio fundamentar a teoria funcionalista da mente contemporânea. Outros pensadores e cientistas da mesma época como Vannevar Bush, também sugeriram, através de construções de máquinas teóricas, como o MEMEX, imaginado por Bush como sendo uma máquina capaz de organizar informações de maneira semelhante à mente humana (Cf. BUSH, 1945), isso nos indica que muitos cientistas, em sua maioria matemáticos ligados a construção da teoria da computação, pensaram na

construção de máquinas funcionalmente equivalentes a mentes humanas, e o computador digital recém criado motiva cada vez mais essa empreitada.

Talvez o experimento mais relevante historicamente para consolidação inteligência artificial foi desenvolvido do “Logical Theorist” em 1955 por Allen Newell e Herbert Simon. O “Logical Theorist” foi um programa de computador desenvolvido para demonstrar teoremas matemáticos. Equipado com um algoritmo de inteligência artificial que lhe permite desenvolver demonstrações acerca de teoremas matemáticos, ou seja, “Logical Theorist” seria um equivalente funcional a mente humana na área de demonstração de problemas matemáticos.

A demonstração de que o Logic Theorist era capaz de provar teoremas foi em si notável. Ele realmente conseguiu provar trinta e oito dos primeiros cinquenta e dois teoremas do capítulo 2 dos Principia . Cerca da metade das provas foi realizada em menos de um minuto cada; a maioria das restantes levou de um a cinco minutos; algumas levaram de quinze a quarenta e cinco minutos; havia forte relação entre o numero de itens da expressão lógica e a duração das provas. Verificou-se que uma prova era mais elegante do que a tentativa de Whitehead e Russell de cinquenta anos atrás – como Simon comunicou a Bertrand Russell, que ficou encantado com esta tirada irônica. (GARDNER, 1995, p. 162).

De acordo com Gardner, “Logic Theorist” corresponde um marco no desenvolvimento da Inteligência artificial, em primeiro lugar pelo impacto que causou na comunidade científica (em especial nos pesquisadores da mente e do cérebro) a época de seu desenvolvimento, em virtude de declarações do próprio Simon como esta: “durante o Natal eu e Allen Newell inventamos uma máquina pensante” (cf. GARDNER, 1995). Em segundo lugar pelas contribuições metodológicas na área da computação que foram essenciais para o desenvolvimento futuro da computação da inteligência artificial, como o processamento em lista que fundamenta a base da linguagem de programação LISP que é até hoje utilizada por pesquisadores de inteligência artificial, e a noção de ferramentas heurísticas necessárias para tratar problemas de explosão combinatória e que atualmente corresponde um campo exclusivo da inteligência artificial. Apesar da polêmica causada pelas declarações fortes de Simon é possível perceber de uma forma prática que a prova do Teorema 2.85 produzida pelo “Logic Theorist” era realmente mais elegante do que a prova produzida laboriosamente à mão por Russell e Whitehead, esse fato, ao mesmo tempo em que espanta muitos pesquisadores de diversas áreas, trás uma nova possibilidade para tratarmos o problema da natureza da mente sob um ponto de vista funcional.

Com o advento de “Logic Theorist” ficou demonstrado que uma máquina pode ser algo mais que um mero “tritador de números”. Ao resolver os teoremas “Logic Theorist” utilizava os mesmos métodos e procedimentos utilizados por agentes humanos, porém algumas vezes o programa optava por uma nova forma, ou uma forma mais elegante do que aquela adotada por agentes humanos realizando a mesma tarefa, assim, ao realizar tarefas que antes eram tidas como exclusivamente “mentais” ou “abstratas”, nada mais normal que admitir que tais computadores estivessem de fato executando uma tarefa semelhante ou similar ao “pensar” humano.

Essa perspectiva funcionalista da mente inaugurada por Turing, que orientou o desenvolvimento tanto do programa “Logic Theorist”, quanto de muitos outros experimentos de inteligência artificial desenvolvidos no período de 1950 a 1970 (cf. TEIXEIRA, 1998 & GARDNER, 1995), foi utilizada por pesquisadores da mente contemporâneos como Daniel Dennett, que não só olhou com bons olhos para a inteligência artificial, como também adotou o critério funcional como delimitador do âmbito mental. Dennett em sua famosa coletânea de artigos “Brainstorms” demonstra que sua teoria da mente é construída mediante uma redução do mental ao funcional, assim, para Dennett, estados mentais seriam apenas funções que são executadas pelo nosso cérebro, da mesma maneira que o bombeamento de sangue é a função exercida pelo nosso coração. Dessa forma, para ele, o cérebro não é o que garante a existência do mental, o que define os estados mentais é a função que eles desempenham. Com isso Dennett não afirma que o conhecimento fisiológico a respeito do cérebro é irrelevante ou deve ser dispensado, mas que devemos estudar o cérebro em seu aspecto físico (ou fisiológico) tão somente para esclarecer a função que ele desempenha. Um dos conceitos centrais da teoria da mente de Dennett é o conceito de sistemas intencionais, de acordo com Dennett, existem três níveis de explicação para um sistema complexo: o nível físico, o nível do design ou planejamento e o nível intencional.

O nível físico (ou perspectiva física) consiste na aplicação das leis da natureza sobre o estado físico de um objeto particular. A limitação da perspectiva física se torna evidente quando, em determinados sistemas, o número de variáveis na sua constituição física é tão grande que não poderia ser capturado por nenhum método conhecido. O nível (ou perspectiva) do design corresponde à descrição da arquitetura desse sistema e como as peças, que se ligam umas às outras, permitem que ele funcione dessa ou daquela maneira. O terceiro nível – o intencional – corresponde à descrição que fazemos do “comportamento” do sistema, identificando neste algum tipo de

racionalidade que nos permita atribuir algum tipo de predicado mental como, por exemplo, “ser inteligente”. (TEIXEIRA, 2008, p. 43).

João de Fernandes Teixeira em sua obra “A mente segundo Dennett” nos apresenta uma descrição sucinta e objetiva dos três níveis de explicação utilizados por Dennett para explicar qualquer sistema complexo, dessa forma, o comportamento humano, o comportamento animal, ou mesmo o comportamento de um computador construído (programado com recursos de inteligência artificial) para jogar xadrez, pode ser explicado utilizando qualquer um dos níveis mencionados por ele. Considerando os três níveis sugeridos por Dennett é possível verificar que a mente não pode ser tão facilmente explicada utilizando o nível físico ou o nível de design, pois sabemos (talvez seja só isso que sabemos com certa tranquilidade) que nosso cérebro é um órgão altamente complexo tanto do ponto de vista de sua constituição física, quanto do ponto de vista do design ou projeto. Por isso mesmo, Dennett não propõe nenhum tipo de redução da mente a qualquer tipo de substância física específica, e sua opção por uma abordagem funcional da mente é demonstrada através do nível de explicação intencional, no qual atribuímos certa racionalidade (na forma de intenções, crenças ou desejos) a determinado sistema complexo no sentido entender ou prever seu comportamento. Assim, o nível de explicação intencional, nos permite prever o comportamento de um sistema complexo atribuindo a esse sistema intenções, crenças e desejos. Um exemplo utilizado por Dennett é a utilização do nível intencional para explicar o comportamento de um computador que joga xadrez.

Os melhores computadores que jogam xadrez atualmente são praticamente inacessíveis à predição segundo a postura de projeto ou a postura física. Eles se tornaram muito complexos mesmo para seus próprios projetistas os verem segundo a postura de projeto. A melhor esperança de um homem para derrotar uma máquina desse tipo em um jogo de xadrez consiste em prever suas respostas descobrindo qual o melhor movimento, ou o movimento mais racional, dada as regras e objetivos do xadrez.(DENNETT, 2006, p. 36).

Dennett não só diz que a o nível intencional é o melhor possível, mas sim que é o único possível, para explicar e prever o comportamento de um computador que joga xadrez. Neste caso, isso quer não dizer que tal máquina tenha de fato uma mente, ou seja capaz de se sentir feliz ou triste, mas sim que, essa máquina pode ser considerada funcionalmente equivalente a mente humana ao jogar xadrez. Através dessa correspondência funcional entre mentes e

máquinas verificamos que a postura intencional utilizada por Dennett é similar, ou no mínimo compatível, com o ponto de vista de Turing, expresso através do jogo da imitação, ambos pensadores optaram por uma descrição da mente em termos da função que ela desempenha. Dennett, assim como Turing em “computação e inteligência”, entende que não existe diferença essencial entre uma inteligência “natural” (oriunda das mentes humanas) e uma inteligência artificial (oriunda das máquinas), como sua caracterização da mente é funcional, a discriminação entre os substratos que as realizam não pode ser justificada. Assim a inteligência artificial pode ser “não natural” na medida em que não corresponde ao resultado da evolução biológica que gerou a mente humana, mas se o critério funcional é legítimo para uma definição da mente, a inteligência artificial pode se mostrar uma excelente ferramenta para compreensão de diversos sistemas complexos dos quais a mente é apenas um caso particular.

2.2 – A quebra do paradigma Representacionalista na Inteligência Artificial

Outros sistemas dotados de I.A. simbólica foram desenvolvidos nessa mesma época, como é o caso de “ELIZA”¹² criado em 1966, por Joseph Weizenbaum, com o objetivo de simular a conversa entre paciente e psicólogo, e também do software “SHRDLU” desenvolvido por Terry Winograd em 1970. O sucesso e otimismo inicial da inteligência artificial começam a ser abalados a partir do desenvolvimento de softwares como “SHDLU”, que nos servirá de exemplo para demonstrar alguns dos principais motivos que levaram a Inteligência artificial, pautada no modelo de representação do conhecimento, a quebrar alguns paradigmas com relação à construção de seus softwares e hardwares.

O que SHRDLU faz ? Resumidamente, ele responde perguntas sobre o seu mundo de blocos feitas por um entrevistador, executa comandos sobre os objetos simulados do mundo de blocos e relata os resultados. Às vezes são ações simples como *pegar*, *mover* ou *colocar*; enquanto outras vezes são ações em um nível mais alto de abstração como *pegar*, *mover* ou *colocar*. (GARDNER, 1995, p. 175).

SHRDLU foi projetado para operar dentro de um mundo fictício de blocos, dentro desse mundo de blocos existem vários objetos (CUBOS, PIRAMIDES E ETC) sendo que

¹² Para maiores informações sobre ELIZA (ver GARDNER, 1995, p. 171 e 172)

“SHRDLU” pode executar ações e responder perguntas de um entrevistador com relação a esses objetos. Apresentaremos abaixo um diálogo, onde as respostas em MAIÚSCULAS correspondem a uma resposta ou ação executada por “SHRDLU”.

Você tocou em alguma pirâmide antes de colocar a pirâmide verde sobre o cubo pequeno?

SIM, A VERDE.

Quando você a pegou ?

QUANDO EU ESTAVA EMPILHANDO O CUBO VERMELHO, UM GRANDE BLOCO VERMELHO E UM GRANDE CUBO VERDE.

Por quê ?

PARA ME LIVRAR DELA..

Por que você fez isso?.

PARA LIBERAR O CUBO VERMELHO.

Por que você LIBEROU esse CUBO?

PARA COLOCÁ-LO SOBRE UM GRANDE CUBO VERDE.

Por que você fez isso?.

POR QUE VOCÊ ME PEDIU PRA FAZER.(DENNETT, 2006, p. 170 e 171)

Um observador que não conhece as regras de funcionamento de “SHRDLU” poderia ficar espantado ao observar suas peripécias no mundo de blocos e suas respostas humanamente sensíveis, porém para um observador que tenha um mínimo conhecimento em computação, “SHRDLU” poderia parecer uma grande enganação na medida em que suas respostas já estão pré-programadas, de alguma forma, em seu código, e foram elaboradas justamente com toques de “sensibilidade humana” para causar exatamente esse efeito de espanto naquele que observa seu funcionamento.

Apesar de softwares como “SHRDLU” e “ELIZA” terem sido objeto de críticas vigorosas, como aquela desenvolvida por John Searle, em seu célebre artigo “Mentes, Cérebros e Programas” publicado em 1981. Tais experimentos na verdade não tinham um objetivo de reproduzir ou explicar o funcionamento da mente humana em todos seus aspectos, pelo contrário as respostas embutidas na máquina era uma forma dos desenvolvedores de tais softwares se divertirem ao observar o espanto de observadores desinformados. Conforme

salienta Daniel Dennett, ao comentar os feitos e efeitos causados pelo de fenômeno “SHRDLU”.

A comunidade de IA paga um preço por essa brincadeira mal compreendida e fascinante, não apenas ao contribuir para a imagens das pessoas da IA como embusteiros e picaretas, mas ao alimentar más concepções mais sérias sobre o ponto de pesquisa da IA. (DENNETT, 2006, p. 170 e 171).

“SHRDLU” é um software que foi desenvolvido em 1970, mas que ainda apresenta uma das principais dificuldades encontradas no campo da I.A., assim como posteriormente no campo da robótica. O problema principal está relacionado ao tipo de representação que deve ser utilizado em uma máquina que pretende demonstrar aspectos cognitivos de alto nível como a capacidade de reconhecer e organizar objetos. A representação que era utilizada por “SHRDLU” correspondia a um mundo muito simples, um mundo que poderia ser representado de maneira completa, alias, acreditava-se que a chave para o sucesso em I.A. seria uma boa representação do mundo, e uma boa representação do mundo deveria conter todas as informações disponíveis nele. Assim, com todas as informações do mundo disponíveis era necessário apenas um mecanismo central de análise, que com base em um conjunto de regras iria decidir a ação adequada. Esse tipo de representação completa do mundo (mundo de blocos) se tornou popular no ambiente de pesquisas relacionado à inteligência artificial e robótica.

Nos final dos anos sessenta e início dos anos setenta, o mundo de blocos (block world) tornou-se um domínio bem popular para a pesquisa em IA, pois sugeria uma semântica uniforme e simples. A chave para o sucesso era representar o estado do mundo completamente e explicitamente. Técnicas de busca poderiam, então, ser usadas para a criação de planos dentro deste mundo tão bem compreendido.(BROOKS,1991, p.145)

Graças ao desenvolvimento de softwares como “SHRDLU” foi possível perceber de forma prática que o processo de construção máquinas capazes de interagirem com o mundo real não seria muito mais difícil do que se pensava. O esquema de representação de “SHRDLU” foi construído com base em um exercício de introspecção de seu programador, ou seja, Winograd neste caso formulou todas as regras para todos os comportamentos possíveis de SHRDLU o que conseqüentemente limitou suas capacidades. Em outras palavras, poderíamos afirmar que SHRDLU utiliza e depende de um modelo de representação (baseado

em regras e símbolos) para agir, porém o software em si, não possui a capacidade de criar ou gerar tais regras que orientem o seu agir, fazendo com que a dependência de algum tipo de representação pré-programada limite o seu comportamento e conseqüentemente restrinja qualquer possibilidade de aprendizado. O mecanismo original que possibilitou uma tomada de decisão inteligente está na verdade na cabeça do programador e não no software em si, com sorte se a representação utilizada pelo programador estiver compatível com a situação na qual o programa está inserido, o mesmo irá apresentar um comportamento inteligente. Brooks ressalta que esse exercício de abstração realizado na construção das representações do mundo é o ponto central da inteligência.

Mas esta abstração é a essência da inteligência, e justamente o núcleo do problema a ser solucionado. No esquema atual, a abstração é feita pelos pesquisadores, de modo que sobra pouco a fazer para os programas de IA, exceto buscar.(BROOKS,1991,p. 146)

O modelo tradicional (utilizado para construção de “SHRDLU”) faz com que os programadores façam todo o trabalho “inteligente” pelas máquinas, não restando espaço para que elas mesmas possam demonstrar algum comportamento realmente inteligente. O mundo de blocos é uma tentativa de representar o mundo real (em sua totalidade), porém esse processo se mostrou um tanto quanto difícil e complexo quando levamos em consideração a grande quantidade de elementos que precisam ser representados e a falta de padrões que encontramos em um ambiente real.

O mundo de blocos foi usado até para pesquisas sobre visão e robótica móvel, pois oferecia fortes restrições sobre o processamento perceptual necessário [12]. Eventualmente, foram feitas críticas no sentido de que o mundo de blocos era um “mundo de brinquedo” e que, dentro dele, havia soluções simples de propósito geral para o que se poderia considerar problemas mais gerais. (BROOKS,1991,p. 145)

Percebemos então dois gargalos principais no modelo tradicional da inteligência artificial, o primeiro seria a necessidade colocar nas máquinas algum mecanismo que fizesse com que ela tome decisões por si mesma, de forma que seu comportamento seja até certo ponto imprevisível. O Segundo ocorre justamente em virtude do primeiro, a falta de capacidade de tomada de decisões, desde o nível mais baixo (desviar de um obstáculo) até o nível mais alto (responder uma pergunta), faz com que exista a necessidade de uma

representação muito complexa, impossível de se construir, uma vez que o mundo real é dinâmico e conta com um número infinito de elementos.

Assim à medida que novos robôs móveis foram sendo desenvolvidos, maior foi à necessidade de mudança de paradigma, pois o software ao ganhar um “corpo”, precisaria imediatamente dos mecanismos de abstração próprios, necessários para que ele mesmo pudesse agir e interagir com o mundo real.

Na tentativa prover tais máquinas com mecanismos próprios de representação várias linhas de pesquisa¹³ dentro da Inteligência Artificial foram desenvolvidas, dentre elas iremos destacar neste artigo a estratégia utilizada por Rodney Brooks que visa resolver, ou pelo menos apontar um caminho para solução dos problemas encontrados em dotar um robô móvel, ou máquina do gênero, com um sistema robusto que permite sua interação com o mundo real.

2.2 – A Nova Abordagem

Enquanto a abordagem tradicional trazia consigo o slogan “A boa representação é a chave para a I.A.”, alguns pesquisadores pareciam caminhar em uma direção contrária, como é o caso do pesquisador americano Rodney Brooks, que desenvolve uma crítica ao modelo representacionalista utilizado até então.

A pesquisa em inteligência artificial encontra-se detida pela questão da representação. Quando a inteligência é abordada de maneira incremental, sendo então compreendida estritamente como responsável pela criação de uma interface com o mundo real através da percepção e da ação, sua dependência para com a representação desaparece.(BROOKS,1991,p.139)

A estratégia utilizada por Brooks rompe totalmente com a abordagem tradicional, segundo ele “A representação é uma unidade de abstração errada quando se trata de construir as partes mais robustas dos sistemas inteligentes.”(1991a. p.141). Ele percebe que a metodologia proposta pela abordagem tradicional é uma construção de cima para baixo (*top-down*), ou seja ela busca a execução de tarefas que exigem alto nível cognitivo,

¹³ Existem uma série de linhas de pesquisa dentro da I.A. como o *aprendizado de máquinas, vida artificial, algoritmos genéticos, entre outras*, que trabalham de certa maneira com algum tipo representação baseada no comportamento emergente, ou seja, não exigem a representação. Para maiores informações ver (COPIN, 2004) capítulos 10,13 e 14, e (TEIXEIRA,1998) segunda e terceira parte.

utilizando representações que acabam por prejudicar a autonomia da máquina em relação a sua interação com o mundo. Assim ele propõe uma construção de baixo pra cima, cujo objetivo principal é dotar a máquina com os mecanismos necessários para execução de comportamentos simples que não requerem a existência de qualquer tipo de representação interna do mundo.

A estratégia de Brooks será bottom-up (de baixo para cima): a simulação do comportamento inteligente deve ter como ponto de partida os comportamentos simples, mundanos, que não requerem a existência prévia de representações. (TEIXEIRA, 1998,p. 134)

Brooks propõe uma nova metodologia para construção de robôs que exibam comportamentos inteligentes sem que haja necessidade de nenhum tipo de representação, a qual ele nomeia de “arquitetura de subsunção”. Encontramos uma ótima definição resumida da referida arquitetura em Copin:

A arquitetura de subsunção é uma arquitetura em camadas que foi projetada para implementar robôs físicos e que não envolve qualquer inteligência centralizada ou mecanismo de controle. O agente nessa arquitetura tem um conjunto de entradas, um conjunto de ações possíveis e um conjunto de módulos em camadas, cada um dos quais projetados para controlar algum aspecto do comportamento do agente. Cada camada é capaz de inibir o comportamento das camadas abaixo dela. (COPIN, 2004, p. 483)

O objetivo de Brooks é uma construção de um sistema inteligente e reativo partindo de baixo para cima, construindo as partes mais básicas que correspondem aos mecanismos de orientação do robô no mundo real, sua programação não deve conter representações explícitas do mundo, mas sim representações simples decompostas em camadas produtoras de atividades, tais camadas (*layers*) seriam então responsáveis pela interação do robô com o mundo.

Cada sistema de produção de atividade ou comportamento conecta individualmente uma sensação a uma ação. Nós nos referimos a aos sistemas de produção de atividade como sendo camadas. Uma atividade é um padrão de interações com o mundo. (...). Escolhemos a palavra atividade, entretanto, porque nossas camadas precisam decidir quanto agir por si mesmas; não se tratam de sub-rotinas que estão dependentes de alguma outra camada e que são evocadas por elas. (BROOKS,1991,p.151)

Alguns agentes autônomos ou robôs foram desenvolvidos no MIT (ver TEIXEIRA, 1998, p. 138-141) sob essa perspectiva, a título de exemplo, apresentaremos resumidamente o funcionamento de um dos primeiros deles, batizado sob o nome de “Allen”. Allen é equipado com alguns sonares¹⁴ (que funcionam como seu aparelho cognitivo) e possui três camadas produtoras de atividade, a primeira é responsável por evitar obstáculos, a segunda faz com que ele se mova aleatoriamente a cada 10 segundos, e a terceira faz com que ele identifique lugares distantes e vá em direção a eles. Cada camada produz suas atividades de forma independente, de forma que, quando “Allen” visita lugares distantes (3ª camada) e encontra algum obstáculo no caminho, ele consegue se desviar (1ª camada) sem que haja uma representação explícita do mundo para tanto.

Na nova abordagem (bottom) a decomposição é feita em termos de módulos de geração de comportamento sendo que cada um dos quais conecta a sensação a ação. Camadas são adicionadas de forma incremental, e novas camadas podem depender de camadas anteriores que operem com sucesso, mas não as chame de sub-rotinas explícitas.(BROOKS, 1991c. p. 1229)

De acordo com Brooks poderíamos obter um sucesso maior na construção de sistemas inteligentes observando a própria natureza, assim, perceberíamos que sistemas complexos podem ser construídos forma incremental. E ainda, o quão importante são os testes práticos realizados a cada passo do desenvolvimento, de forma análoga ao processo de evolução por seleção natural. Segundo ele, seria mais conveniente alcançar a inteligência em um nível mais alto (por exemplo compreensão da linguagem natural) a partir do desenvolvimento de uma inteligência mais simples que surge a partir da própria interação da máquina com o mundo real.

A competência comportamental do sistema é melhorada ao adicionar uma rede mais específica de comportamento a uma rede existente. Esse processo é chamado de *layering*. Esta é uma analogia crua e simplista do desenvolvimento evolucionário. Assim como a evolução, a cada estágio do desenvolvimento os sistemas são testados. Cada uma das camadas é uma peça de produção comportamental da rede a sua própria maneira, embora ela possa implicitamente confiar na presença de partes anteriores de rede. Por exemplo, uma camada de exploração não precisa

¹⁴ O sonar é um instrumento inicialmente usado em época de guerra para a localização de submarinos, seu funcionamento consiste no envio e recebimento de ondas de som, alguns animais como o golfinho e o morcego utilizam desse modelo como ferramenta de orientação no mundo.

explicitamente evitar obstáculos, já que o desenhista sabe que uma camada existente vai cuidar dele. Um esquema arbitrário de prioridade fixa é usado para lidar com os conflitos. (BROOKS, 1991c. p. 1229)

Este processo é chamado de layering e constitui, na verdade, uma analogia simplista com o processo evolucionário. Desta concepção emerge a idéia de que o “critério de decomposição” de um ser inteligente não é por módulos funcionais e sim por módulos de atividade: o ponto de partida da cognição é a ação e não a representação. (TEIXEIRA, 1998, p. 137)

Brooks ressalta que devemos procurar percorrer o mesmo caminho trilhado por nos seres humanos ao longo de nosso desenvolvimento histórico evolutivo, procurar representar o conhecimento em sua totalidade, ou mesmo criar máquinas capazes de compreender a linguagem natural antes de alcançar uma inteligência de um nível mais baixo, ou mais simples, seria impossível.

A evolução demorou 3 bilhões de anos para sair das células individuais para os insetos, e mais 500 milhões de anos para sair daí para os seres humanos. Não citamos estes dados para fazer estimativas sobre nossa performance futura, mas para indicar como o nível de inteligência de inseto não é algo trivial. (BROOKS, 1991, p. 164)

Nossa crença é que os tipos de camadas de controle produtoras de atividade que estamos desenvolvendo (tarefas relacionadas à mobilidade, visão e sobrevivência) são pré-requisitos necessários para o nível mais alto de inteligência que atribuímos aos seres humanos. (BROOKS, 1991, p. 164)

O que se percebe nesse movimento é uma mudança na própria forma de criar sistemas artificiais, a boa representação do mundo como uma solução para produção de comportamento inteligente não se mostrou tão eficaz, e deu lugar a uma tentativa de construção de um sistema cognitivo artificial que não depende de representação alguma, seu comportamento inteligente é apenas um reflexo baseado na própria dinâmica do ambiente no qual ele se encontra (mundo real).

Essa mudança na forma de criar sistemas artificiais não está relacionada apenas com inteligência artificial enquanto disciplina ligada a Ciência da Computação, uma vez que o paradigma representacionista da mente está presente também nas teorias Funcionalistas da mente, conforme foi visto nos capítulos anteriores a abordagem tradicional de Inteligência artificial, aplicada a experimentos como ELIZA, SHRDLU, Logical Theorist entre outros é

totalmente representacionista. Dessa forma, partindo do pressuposto de que tais experimentos de Inteligência Artificial são substrato físicos para a “comprovação” ou ao menos a “verificação” da teoria funcionalista da mente, a quebra de paradigma representa uma grande contribuição para o desenvolvimento da teoria funcionalista da mente uma vez que os experimentos realizados, como o SHRDLU, demonstram a limitação que existe dentro da própria teoria funcionalista da mente, e ainda, a nova abordagem proposta por Brooks representa também uma grande contribuição para elaboração de uma teoria da mente que supera os problemas encontrados dentro do paradigma representacionista da mente vigente em todos os experimentos de Inteligência Artificial desenvolvidos até então.

CAPÍTULO 3 - CONSIDERAÇÕES FINAIS

Conforme foi demonstrado no capítulo 1, através de uma breve apresentação do panorama geral da filosofia da mente, encontramos um conjunto de teorias, elaboradas desde os tempos mais antigos, ainda na filosofia antiga com Platão, passando pela filosofia moderna com Descartes e outros pensadores, até a filosofia contemporânea. O esforço empreendido na elaboração das teorias da mente, em um sentido geral, sempre está ligado a elucidação (epistêmica e ontológica) de conceitos como Mente, Consciência, Raciocínio, Pensamento.

No capítulo 2, nas seções 2.1 e 2.2, foram apresentados argumentos que demonstram como a inteligência artificial colaborou para a construção e amadurecimento da filosofia da mente, de maneira geral, e da teoria funcionalista da mente, de maneira específica. Assim como foi mostrada a relação existente entre o Funcionalismo e a Inteligência artificial.

A partir de Turing, a inteligência artificial carece da filosofia da mente para fundamentar e orientar a modelagem de seus sistemas, ao passo que a filosofia da mente se apropria da perspectiva funcionalista da mente inaugurada por Turing. Atualmente, filósofos da mente, como Daniel Dennett têm defendido que essa aproximação entre essas duas disciplinas é fundamental para o desenvolvimento de ambas.

Os filósofos, tenho dito, deveriam estudar IA. Os pesquisadores de IA deveriam estudar filosofia? Sim, a menos que eles estejam contentes em reinventar a roda a cada poucos dias. Quando a IA reinventa a roda, ela é caracteristicamente quadrada, ou, no melhor dos casos, hexagonal, e pode apenas fazer algumas centenas de revoluções antes de parar. As rodas do filósofo, por sua vez, são círculos perfeitos, e não requerem, em princípio, nenhuma lubrificação; e podem ir pelo menos em duas direções ao mesmo tempo. (DENNETT,2006, p. 183)

Acreditamos que a relação existente entre a inteligência artificial e a teoria funcionalista da mente se dá em um nível de dependência, pois o desenvolvimento da inteligência artificial depende de um ponto de vista funcional acerca do mental, basicamente é só isso que importa pra ela, pois seu compromisso com a construção de máquinas funcionalmente equivalentes a mente humana, por outro lado, temos que o principal objetivo da teoria funcionalista da mente é

uma explicação que dê conta da natureza da mente, levando em consideração única e exclusivamente a função que a mente desempenha no mundo, nesse sentido seus objetivos se sobrepõe e como disse DENNETT “um encontro de mentes é necessário” (2006, p. 183). Uma das justificativas para essa aproximação, segundo DENNETT é que “a IA compartilha com a filosofia (em particular, com a epistemologia e a filosofia da mente) o estatuto da investigação mais abstrata dos princípios da psicologia” (2006, p.167). Assim, entendemos que um diálogo de aproximação entre essas disciplinas pode ser interessante para ambos os lados, uma vez que em determinados pontos se persegue os mesmos objetivos.

Mais adiante, nas seções 2.2 e 2.3, foram apresentadas respectivamente as teses relacionadas a quebra do paradigma Representacionalista na inteligência artificial e a nova abordagem, proposta por Rodney Brooks, no sentido de demonstrar que a Inteligência artificial, enquanto modelo que possibilita a verificação de teorias funcionalistas da mente, deve ser revisto. Assim, enquanto a Inteligência artificial aponta para problemas práticos que ocorrem no seu desenvolvimento, ela aponta também para problemas que são encontrados no cerne da teoria funcionalista da mente. Uma nova abordagem que abdica da representação para a simulação do comportamento inteligente só tem a engrandecer o campo de estudo teórico que serve de base para a teoria funcionalista da mente.

CONCLUSÃO

Durante o processo de pesquisa que culminou neste trabalho foram analisados textos de diversos autores e sobre diversas teorias da mente com o objetivo principal de entender qual é o arcabouço geral da filosofia da mente, através de uma breve análise de várias teorias, para que mais adiante fosse possível entender qual o cenário em que a teoria funcionalista da mente nasceu e como ocorreu o seu desenvolvimento até os dias de hoje, para tanto várias questões foram abordadas, como a criação do computador digital e o desenvolvimento dos primeiros experimentos de inteligência artificial. Nesse ínterim percebemos que a partir dos trabalhos de Alan Turing, principalmente no que tange a elaboração do critério funcional como delimitador do âmbito mental, foi possível não só a consolidação de um substrato físico capaz de testar e avaliar pressupostos da teoria funcionalista da mente, mas também de um substrato teórico passível de pesquisa pelos pesquisadores ligados ao funcionalismo. A partir da análise do desenvolvimento da Inteligência artificial, do estudo dos experimentos realizados por Allen Newell e Herbert Simon foi possível perceber a consolidação do modelo representacionalista inaugurado por Alan Turing, porém com o desenvolvimento dos computadores e com o aperfeiçoamento das técnicas de programação o modelo representacionalista chega a um patamar que precisa ser transposto, e através da leitura dos textos do pesquisador de Inteligência artificial Rodney Brooks foi possível perceber que as questões tratadas sob a égide da construção de uma nova abordagem para a robótica tem muito a acrescentar a teoria funcionalista da mente de maneira específica e a filosofia da mente de maneira geral.

REFERÊNCIAS BIBLIOGRÁFICAS

COPPIN, Ben. Inteligência Artificial. 1ª Edição. Rio de Janeiro: LTC, 2004.

CHURCHLAND, P.M. Matéria e Consciência: uma introdução contemporânea a filosofia da mente. São Paulo: UNESP, 2004

BENTLEY, Peter. Biologia Digital: como a natureza está transformando nossa tecnologia e nossas vidas. São Paulo: Berkeley Brasil, 2002.

BROOKS, R. Inteligência sem representação. Tradução: Pedro R. de Oliveira. filosofiadamente.org, (Acesso em 12 de junho de 2012)

_____, R.A.. Intelligence without reason – Proceedings of the 12th. IJCAI pp. 565-95, San Mateo, CA: Morgan Kauffmann. 1991b

_____, R. A., "New Approaches to Robotics", Science (253), September 1991c, pp. 1227–1232.

_____, R.A., "The Relationship Between Matter and Life", Nature, Vol. 409, January 18, 2001, pp. 409–411. (Invited)

DENNET, Daniel. Brainstorms: escritos filosóficos sobre a mente e a psicologia. Tradução Luiz Henrique de Araújo Dutra. São Paulo: Editora UNESP, 2006

_____, Daniel. Hal cometeu um assassinato?. Tradução Pedro Rocha de Oliveira. 1996, Disponível em: <http://www.filosofiadamente.org/images/stories/textos>

DESCARTES, R. Meditações da Filosofia Primeira. Tradução de J. Guinsburg e Bento Prado Júnior. São Paulo: Abril, 1973. (Col. Os Pensadores).

_____, O Discurso do Método. Tradução de Ciro Mioranza. São Paulo: Escala, 2009

DONATELLI, M. C. de O. F. Conarius e memória na carta de 1 de abril de 1640 de Descartes a Mersenne Sci. stud. vol.1 no.1 São Paulo Mar. 2003

DREYFUS, Hubert. What Computers Still Can't Do: A Critique of Artificial Reason. Nova York: Harper & Row, 1972.

GARDNER, H. A Nova Ciência da Mente. São Paulo: EdUSP, 1996.

HAUGELAND, John. Máquinas semânticas: uma introdução ao design de mentes Tradução: Vítor Guerreiro. 1981, Disponível em: <http://mlag.up.pt/wp-content/uploads/2011/05/HAUGELAND.pdf>

LIMA, M. L.F. I Naturalismo Biológico: a solução dualista de John Searle para o problema mente-corpo. 2010. Dissertação (Mestrado) – Universidade Federal do Ceará, Fortaleza 2010

JACKSON, Frank. “Epiphenomenal Qualia”. *Philosophical Quarterly*, 32, pp. 127-36, 1982.

LEAL-TOLEDO, G. O Argumento dos Zumbis na Filosofia da Mente: são zumbis físicos logicamente possíveis?. 2005. Dissertação (Mestrado) Instituto de Filosofia e Ciências Humanas, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2005

_____, Dennett e Chalmers: Argumentos e intuição. *Trans/Form/Ação*, São Paulo, 29(2): 123-132, 2006

MACHADO Nivaldo; LEAL-TOLEDO, Gustavo. Entre o filósofo e o cientista: poderá a máquina sentir saudade ?. Rio do Sul: UNIDAVI-PROPPEX, 2012.

MOORE, J;NEWELL, Allen, "How can Merlin understand?" (1973). *Computer Science Department*. Paper 1613.

NAGEL, T. What Is It Like to Be a Bat? In: *Philosophical Review*, v. LXXXIII, n. 4, 1974, p. 435 – 450

PORTO, Leonardo Sartori. Uma investigação filosófica sobre a inteligência artificial. *Informática na Educação: teoria & prática*, Porto Alegre. V.8, n. , p. 11-26, jan./jun. 2006.

SEARLE, J.R. *A Redescoberta da Mente*. São Paulo: Martins Fontes, 1997.

_____, “Mentes, cérebros e programas”. Tradução de Cléa Regina de Oliveira Ribeiro. In: TEIXEIRA, João de Fernandes (Org.). *Cérebros, Máquinas e Consciência: uma introdução à Filosofia da Mente*. São Carlos: Editora da UFSCar, 1996.

_____, *Consciência e Linguagem*. São Paulo: Martins Fontes, 2010.

_____, *Intencionalidade*. São Paulo: Martins Fontes, 2002.

_____, *Mente, Cérebro e Ciência*. São Paulo: Edições 70, 1984.

_____, *O Mistério da Consciência*. São Paulo: Ed. Paz e Terra, 1998. TEIXEIRA, J. F. *A mente segundo Dennett*. São Paulo: Perspectiva S.A., 2008

_____,Mentes e Maquinas. Porto Alegre: Artes Médicas, 1998

_____,Mente, cérebro e cognição. Petrópolis, RJ: Vozes,2000

TURING, Alan. Computação e inteligência. Tradução de Fábio de Carvalho Hansen. In: TEIXEIRA, João de Fernandes (Org.). Cérebros, máquinas e consciência: uma introdução à Filosofia da Mente. São Carlos: Editora da UFSCar,1996.

_____, Alan, 1946, ACE Reports of 1946 and Other Papers.In: DENNETT, Daniel. Hal cometeu um assassinato?. Tradução Pedro Rocha de Oliveira. 1996,url: <http://www.filosofiadamente.org/images/stories/textos>

_____, Alan. “On computable numbers, with an application to the Entscheidungsproblem” .1936.Disponível em: <http://www.dcc.ufrj.br/~luisms/turing/>

TEIXEIRA, J. F. A mente segundo Dennett. São Paulo: Perspectiva S.A.,2008

_____,Mentes e Maquinas. Porto Alegre: Artes Médicas, 1998

_____,Mente, cérebro e cognição. Petrópolis, RJ: Vozes,2000

VASCONCELLOS, S.J.L. Filosofia da mente: uma revisão crítica. PSICO, Porto Alegre, v. 38, n. 2, PP 190-195, maio-ago. 2007.

VON NEUMAN, John. The Computer and the Brain. New Haven, Conn., Yale University Press. 1958.