



EVOLUÇÃO, FUNCIONALISMO E INTENCIONALIDADE

André Rosolem Sant'Anna¹

Resumo: O problema da intencionalidade em filosofia da mente diz respeito ao modo pelo qual a mente humana e de outros animais estão relacionadas ao mundo, o que exige uma análise cuidadosa por parte de um projeto filosófico que pretenda explicar de forma satisfatória este tópico. John Searle, em *Minds, Brains and Programs*, oferece uma crítica contundente às tentativas de se explicar a intencionalidade e a atribuição de significado aos estados mentais a partir da perspectiva funcionalista a qual ele classifica por “Inteligência Artificial Forte”. Na visão de Searle, a produção sintática das relações lógicas do cérebro humano não é suficiente para reproduzir um sistema com intencionalidade e significado. Em contraposição a Searle, Daniel Dennett argumenta em favor da tese de que sistemas inteligentes e aparentemente designados podem ter sua origem em processos simples e mecânicos como os processos algorítmicos, o que, em princípio, parece indicar a possibilidade de uma abordagem da intencionalidade nestes mesmos pressupostos. Tendo em vista a proposta de Dennett, este artigo tem como objetivo expor alguns aspectos da teoria da intencionalidade de Dennett e explicitar algumas considerações baseadas na teoria da evolução que permitem a conciliação de uma teoria funcionalista da mente com o fenômeno da intencionalidade.

Palavras-chave: evolução; funcionalismo; intencionalidade; significado; Dennett.

¹ Graduando em Filosofia pela Universidade Estadual de Maringá (UEM). e-mail: rosolemandre@gmail.com

O PROBLEMA DA INTENCIONALIDADE

O vocabulário mental tradicional no qual crenças, desejos, intenções, medo, amor etc. são utilizados para descrever nossos estados mentais e, de certa forma, explicar nosso comportamento, são tomados por alguns filósofos como expressões de entidades realmente existentes no mundo e cujos referentes, portanto, merecem ser objetos de investigação científica quando se pretende realizar um estudo efetivo da mente.

A concepção tradicional dualista da mente assume uma postura realista em relação aos estados mentais de um ser humano. A separação ontológica proposta por Descartes entre mente e corpo pressupõe que as assim denominadas “entidades mentais” estejam situadas em um plano imaterial, ou seja, nossas crenças e desejos, por exemplo, embora existentes neste plano imaterial, não seriam entidades genuínas do mundo físico.

Guiados pelas tentativas materialistas de abordagem do problema mente-corpo, grande parte dos filósofos da mente contemporâneos assumem o projeto de se explicar a ocorrência dos estados mentais acima mencionados dentro de uma ontologia materialista. Neste sentido, parte do projeto científico dedicado a este escopo seria destinado a explicar como ocorrem as instanciações de crenças, desejo etc. em uma estrutura orgânica como o cérebro.

Uma abordagem materialista, na qual a busca de correlatos físicos para a ocorrência de estados mentais se faz necessária, deve ser capaz de explicar em termos suficientemente físicos o fenômeno da *intencionalidade*. A intencionalidade é tomada por muitos filósofos como a característica que permite o direcionamento da mente em relação a objetos no mundo físico, característica esta que nos proporciona a capacidade de representação deste mesmo mundo físico². Neste sentido, se tenho a crença de que o

2 John Searle inicia sua obra *Intencionalidade* com a seguinte declaração: “Poderíamos dizer, a título de formulação preliminar, que a Intencionalidade é aquela propriedade de estados e eventos mentais pela qual estes são dirigidos para, ou acerca de, objetos e estados de coisas no mundo.” (SEARLE, 1984, p. 1)

carro estacionado em frente à minha casa é branco, a minha crença está direcionada a um estado de coisas no mundo, a saber, o carro estacionado em frente à minha casa. Em uma perspectiva geral, estados mentais tais como crenças, medo, desejo etc. são estados intencionais porque estão *direcionados* a certos objetos ordinários do mundo. Para se ter uma crença ou um desejo é preciso que se tenha crença *de* alguma coisa ou desejo *de* alguma coisa. Como aponta John Searle (1984), o *de* é o *de* da intencionalidade, é o direcionamento ou a aproximação da crença ou do desejo para o mundo.

Searle (1984, 1992) diz ser a intencionalidade, tal como descrita acima, uma das características marcantes do mental e que uma teoria científica do cérebro deve ser, neste contexto, capaz de explicar como este último é capaz de produzir intencionalidade.

Em sua obra *Intencionalidade*, Searle baseia sua análise lógica dos estados intencionais em uma distinção entre intencionalidade *intrínseca* e intencionalidade *derivada*. O primeiro caso de intencionalidade diz respeito a uma característica idiossincrática da mente humana e de alguns animais³, isto é, o termo *intrínseco* é utilizado no sentido em que a intencionalidade proveniente da mente humana não é derivada de nenhum outro lugar⁴. Já no segundo caso, o tipo de intencionalidade ao qual Searle se refere é uma espécie secundária de intencionalidade⁵. Objetos como mapas e livros,

³ Adotar-se-á nas páginas subsequentes uma definição restrita da intencionalidade intrínseca, isto é, o uso do termo a partir deste momento será restrito ao caso da mente ou do cérebro humano.

⁴ “[...] intencionalidade intrínseca é um fenômeno que seres humanos e determinados outros animais têm como parte de sua natureza biológica. Não é uma questão de como são tratados, ou como se concebem a si mesmos, ou de que forma preferem descrever-se a si mesmos.” (SEARLE, 1992, p. 118)

⁵ “Todos estados intencionais *intrínsecos*, no sentido que lhes dou, o que significa exatamente que são a coisa real e não apenas algo mais ou menos como a coisa real (*como-se*), ou algo que seja o resultado dos empregos da coisa por outro alguém, ou de suas atitudes em relação a ela (*derivados*).” (SEARLE, 1992, p. 120). Tendo em vista esta passagem, é possível notar que o significado ontológico atribuído por Searle à intencionalidade intrínseca (“coisa real”) é mais forte do que o da intencionalidade derivada (“como-se”). Esta última seria, dentro desta terminologia, uma propriedade

por exemplo, possuem apenas intencionalidade secundária ou derivada, uma vez que um mapa, ao ser tomado isoladamente, isto é, na ausência de um observador com intencionalidade intrínseca, constitui apenas um agregado de símbolos sem *significado*.

Neste contexto, as sentenças de um livro como “O carro é branco” ou “As folhas da árvore são verdes” podem representar, através de um conjunto de símbolos, um determinado estado de coisas do mundo. A representação de determinado estado de coisas do mundo confere a estas sentenças um direcionamento a este mesmo mundo, de modo que, de acordo com a definição dada até aqui, elas podem ser ditas possuidoras de intencionalidade. As palavras que constituem estas sentenças, no entanto, não possuem nenhum conteúdo ou significado por si mesmas. O conteúdo intencional, isto é, o conteúdo representativo destas sentenças tem sua origem na mente humana capaz de produzir, na perspectiva de Searle, significados intrinsecamente. Em outras palavras, livros, mapas, sentenças etc. não possuem significado enquanto um agregado de símbolos impressos em um papel, mas somente o terão mediante a adoção de uma postura interpretativa por parte de um indivíduo com intencionalidade intrínseca, o que torna esta atribuição uma atribuição de caráter metafórico.

A capacidade de produzir significado da mente humana, e, portanto, a sua capacidade de produzir estados intencionais, não pode ser reproduzida apenas por uma reconstrução sintática dos aspectos lógicos das operações cerebrais, tal como se pressupõe em uma simulação computacional da mente, uma vez que uma reprodução sintática tal qual a se observa em sentenças de um livro não possui significado intrinsecamente. Neste sentido, Searle argumenta que a mera reprodução sintática da operação cerebral não é uma característica suficiente para que possamos dizer que um sistema possua uma mente de fato. A capacidade de produzir estados intencionais, por conseguinte, só pode residir no poder causal das estruturas orgânicas das quais o nosso cérebro é constituído:

relacional na medida em que seu estatuto ontológico só é garantido uma vez que algum indivíduo com intencionalidade intrínseca adote determinada postura em relação a ela.

O problema com um simulador do cérebro é que ele estaria simulando aspectos errôneos sobre o cérebro. Na medida em que o simulador simula apenas a estrutura formal da sequência dos estímulos de neurônios nas sinapses, ele não terá simulado o que importa sobre o cérebro, nomeadamente suas *propriedades causais*, sua *habilidade para produzir estados intencionais*⁶. (SEARLE, 1980, p. 364) (Tradução e ênfases minhas)

O posicionamento de Searle parece indicar que, dentro do contexto apresentado acima, a intencionalidade intrínseca existente em estruturas orgânicas tal como os seres humanos e outros animais depende intimamente dos aspectos causais subjacentes ao funcionamento do cérebro destes sistemas, o que, em princípio, inviabilizaria a proposta do estudo da mente como um computador digital seguindo os pressupostos da “Inteligência Artificial Forte”.

O FUNCIONALISMO⁷

Alan Turing, ao publicar seu artigo *Computing Machinery and Intelligence* (1950), propõe a seguinte questão: “Podem as máquinas pensar?” (Tradução minha) (TURING, 1950, p. 441), apresentando então um teste que, segundo Turing, seria capaz de identificar se uma máquina é efetivamente capaz de pensar⁸. O *Teste de Turing*, no qual, em linhas gerais, um juiz A tem contato com uma pessoa B e um computador C apenas por respostas fornecidas por estes últimos às perguntas de A. Sendo B

⁶ A crítica de Searle às tentativas de reprodução sintática dos aspectos lógicos das operações cerebrais será tratada com mais detalhes na próxima seção.

⁷ Cabe ressaltar que o que tratamos aqui por funcionalismo, apesar de estar caracterizado de modo geral, é apenas uma das versões do funcionalismo, isto é, o que entendemos por funcionalismo aqui pode ser compreendido na medida em que consideramos a analogia entre mente-cérebro e *software-hardware*. Estes pontos serão desenvolvidos com maior detalhe ao longo desta seção.

⁸ Turing (1950) não coloca a *questão* nestes termos, mas antes, ele propõe um novo *questionamento* frente à *questão* “Podem as máquinas pensar?”. Segundo Turing, questionar se: “Existem computadores digitais que se sairiam bem no jogo da imitação?” (Trad. minha) (TURING, 1950, p. 448) seria uma questão equivalente à primeira. “Jogo da imitação” é o nome atribuído por Turing ao que ficou conhecido como *Teste de Turing*.

uma mulher, esta deve fornecer informações corretas às perguntas de A, de modo que suas informações moldadas pelas perguntas devem ajudar A a identificar o sexo de B. O computador C, por outro lado, deve ser programado para enganar o juiz A, tentando convencê-lo ser ele uma mulher. Assim, se A, um juiz humano, não conseguir apontar C como sendo um computador, isto implicará que C é capaz de “passar” no jogo da imitação, ou Teste de Turing, sendo C, portanto, conforme aponta Turing ao propor a equivalência das questões (Cf. nota de rodapé 7), capaz de pensar.

No contexto da proposta de Turing surge a abordagem do funcionalismo proposta por Hilary Putnam (1973). Segundo Putnam, os estados mentais são estados funcionais que executam determinada função, não sendo eles ontologicamente físicos ou mentais *per se*. Deste modo, um estado mental é definido de acordo com a execução de um papel funcional entre *inputs* e *outputs* e de acordo com sua relação com outros estados funcionais de um sistema. Putnam, referindo-se ao exemplo da dor, descreve o funcionalismo nos seguintes termos:

Eu devo, em resumo, argumentar que a dor não é um processo do cérebro, no sentido de estados psico-químicos do cérebro (ou mesmo de todo o sistema nervoso), mas [a dor] é completamente outro *tipo* de estado. Eu proponho a hipótese de que a dor, ou o estado de estar com dor, é um estado funcional de um organismo como um todo. (PUTNAM, 1973, p. 75) (Tradução minha)

Dado a proposta de Putnam para se considerar estados mentais como estados funcionais que executam uma determinada *função* dentro de um todo, aduzindo assim às relações de *inputs* e *outputs* da qual estes estados são intermediários, surgem propostas baseadas em modelos computacionais e que defendem a analogia da relação entre mente e cérebro com a relação observada entre um programa computacional, denominado *software*, e a estrutura física do computador a qual este programa está subjugado, o *hardware*. De forma resumida, as características da mente humana poderiam ser reproduzidas através da operação sintática

de símbolos realizada por um *software* computacional. O empreendimento que adota tais pressupostos para se realizar um estudo da mente humana é denominado por Searle (1980) como “Inteligência Artificial Forte”.

Em *Minds, Brains, and Programs*, Searle lança mão do argumento do Quarto Chinês, no qual ele se utiliza da distinção entre intencionalidade intrínseca e derivada para tentar demonstrar a falibilidade do programa de pesquisa ao qual ele denomina de “Inteligência Artificial Forte”. Searle propõe que pensemos em uma pessoa dentro de um quarto realizando a operação transitória e causal entre os *inputs* ambientais, como frases em chinês, e os *outputs* comportamentais, como a emissão de frases em chinês coerentes com os *inputs*, guiada apenas pela mera operação sintática de símbolos e regras previamente estabelecidas. Em outras palavras, a pessoa disporia de um conjunto de regras pré-estabelecidas que ditaria a sua operação dos termos em chinês provenientes dos *inputs* para produzir emissões igualmente em chinês coerentes com os *inputs* recebidos. Searle conclui, a partir destas resoluções, que embora o comportamento apresentado pelo Quarto aparente a existência de um processo interno em que há a compreensão de chinês, uma vez que a relação causal observada entre *inputs* e *outputs* é uma relação coerente entre os aspectos semânticos destes mesmos *inputs* e *outputs*, o próprio processo causal responsável por esta transição não compreende chinês, visto que os processos que se dão no interior do Quarto são apenas processos sintáticos e operacionais guiados por um conjunto de regras anteriormente estabelecidas. Neste sentido, afirma Searle, a mera operação ou habilidade sintática não é condição suficiente para que o significado das emissões em chinês seja estabelecido.

De modo geral, o argumento de Searle parece indicar que uma abordagem funcionalista da mente tal como a proposta pelos desenvolvimentos da Inteligência Artificial Forte, no qual estados mentais são identificados com seus papéis causais entre *inputs* e *outputs*, não seria suficiente para explicar como a atribuição de significado ocorre na mente humana. Deste modo, ainda que uma máquina seja capaz de passar no

Teste de Turing, a sua intencionalidade seria apenas uma intencionalidade derivada, provinda da intencionalidade intrínseca de seus criadores, visto que o significado aparentemente inerente aos estados mentais da mente humana não poderiam emergir de uma mera operação sintática de regras previamente estabelecidas.

A PROPOSTA FUNCIONALISTA E EVOLUCIONISTA DE DENNETT

Em *Darwin's Dangerous Idea*, Daniel Dennett propõe uma espécie de universalização dos aspectos lógicos da teoria da seleção natural, argumentando em favor do tratamento da ideia de Darwin, a seleção natural, sob o prisma do que ele denomina de “algoritmo de substrato neutro”. Um algoritmo, na concepção de Dennett, é “um certo tipo de processo formal que pode – logicamente – dar origem a um determinado resultado toda vez em que for aplicado [*run*] ou instanciado.” (DENNETT, 1995, p. 50) (Tradução minha). Enquanto um processo lógico e formal, o algoritmo da divisão matemática, por exemplo, pode ser instanciado tanto com a utilização de um lápis e uma folha de papel quanto com a utilização de uma calculadora. Neste sentido, portanto, um algoritmo pode ser considerado *neutro* em relação ao *substrato* em que ele é realizado.

Ao apresentar sua defesa da ideia de Darwin, Dennett se compromete com uma abordagem *adaptacionista* em relação ao modo pelo qual se pode explicar a presença de determinadas características biológicas nos seres vivos (Cf. ZAWIDZKI, 2007, p. 125-52)⁹. O adaptacionismo, em poucas palavras, defende que a presença de determinado traço biológico em um organismo é resultado de uma pressão seletiva exercida pelo ambiente, de modo que organismos com características mais aptas a este ambiente seriam selecionados. Em outras palavras, organismos com características mais adaptadas às adversidades do ambiente teriam maior chance de se reproduzir.

O processo de seleção no qual as características mais aptas são

⁹ Ver também ELTON, 2003, p. 220-8.

transmitidas pelos indivíduos mais adaptados aos seus descendentes é repetido inúmeras vezes ao longo da história evolutiva, de modo que a acumulação destas características ao longo do tempo é capaz de dar origem a organismos ou parte de organismos aparentemente designados para determinadas funções. Este processo de acumulação de características é denominado de “Princípio da Acumulação do Design” (Cf. DENNETT, 1995, pp. 68-73).

Tendo em vista esta breve caracterização da evolução, é possível notar que a seleção gradativa de indivíduos com características mais aptas a sobreviver em determinado ambiente é um processo mecânico e sem propósitos ou direcionamento, já que características aparentemente designadas podem ter sua origem explicada a partir dos pressupostos da acumulação do *design* que tem sua origem em um processo de seleção caracteristicamente algorítmico. A seleção natural, em outras palavras, opera de acordo com um algoritmo no qual indivíduos mais aptos a determinado ambiente terão maior aptidão para deixar descendentes.

O caráter lógico da ideia que subjaz ao processo da seleção natural permite que as pressuposições lógicas deste processo tenham uma aplicação generalizada, de modo que características aparentemente designadas para determinado fim ou função podem ser explicadas a partir da instanciação de um processo algorítmico dentro de um período de tempo específico¹⁰.

¹⁰ Dennett (1995) alude a um esquema conceitual ao qual ele chama de “Pirâmide Cósmica” (DENNETT, 1995, p. 64). De forma resumida, a Pirâmide Cósmica seria uma organização hierárquica na qual *Deus* se figura no topo da pirâmide, seguido sucessivamente por *Mente* e *Design*. Neste sentido, para que haja artefatos ou organismos designados para um fim ou uma função específica, é preciso que exista uma *Mente* capaz de produzi-los. Esta *Mente*, por sua vez, não pode ter sua origem senão em algo mais perfeito e inteligente do que ela, de modo que *Deus* seria a figura elementar e o topo desta pirâmide. Ao se prosseguir na hierarquia da pirâmide, abaixo do *Design* situam-se a *Ordem*, o *Caos* e o *Nada*. Dennett argumenta que a ideia de Darwin permite que a organização desta pirâmide seja alterada, de modo que a *Ordem* fornecida por um processo algorítmico tal como a seleção natural, somada ao tempo, seriam capazes de produzir *Design* sem que fosse necessário a presença de uma *Mente* ou de *Deus*. “Mas Darwin sugere uma divisão: Deem-me Ordem, ele diz, e tempo, e eu lhes darei Design.”

Baseado na interpretação algorítmica da seleção natural, Dennett, em *The Intentional Stance*, propõe que imaginemos uma situação em que, por algum motivo, desejemos estarmos vivos no século XXV. Dado que possuíssimos conhecimento tecnológico amplamente avançado, poderíamos construir um robô que abrigasse nosso corpo em seu interior em uma espécie de conserva. Este robô, em função dos objetivos do nosso empreendimento, deve ser capaz de interagir com o ambiente de modo que as possibilidades mais aptas para a sua preservação sejam empreendidas quando requisitadas. Se, por exemplo, outras pessoas estivessem interessadas em sobreviver até o século XXV e para tal criassem robôs para armazenarem seus corpos, o projeto de design do nosso robô teria que ser muito mais sofisticado, visto que, além do ambiente, ele teria que interagir com outros robôs que buscam o mesmo objetivo.

A construção deste robô pode ser empreendida a partir da implementação gradual de processos mecânicos e sem inteligência capazes de lidar com peculiaridades específicas exigidas pelo ambiente e pela interação com outros robôs, de modo que, a partir do Princípio da Acumulação do Design, possamos ter como produto final o robô tal qual desejamos inicialmente¹¹.

(DENNETT, 1995, p. 65) (Tradução minha). A mente humana, neste contexto, uma vez que seja considerada como produto da evolução por seleção natural, estaria sujeita ao mesmo processo de *Design*. Em outras palavras, considerando a inversão de Darwin da ordem da Pirâmide Cósmica, a mente humana seria o resultado da instanciação da *Ordem* através do processo algorítmico dentro de um período de tempo específico. Este é um dos pontos principais do projeto de naturalização da mente de Dennett que será desenvolvido nas linhas subseqüentes.

¹¹ Dennett (1987) explicita com detalhes este processo utilizando o exemplo de um termostato ao qual diferentes estruturas formais são adicionadas de modo que sua sofisticação semântica aumenta gradativamente: “Um modo mais formal de se dizer isto [a atribuição gradativa de estruturas formais] é que a classe indistinguível de modelos satisfatórios do sistema formal corporificada nos seus estados internos torna-se cada vez menor na medida em que nós adicionamos estas complexidades; quanto mais adicionamos, mais rica, mais exigente ou mais específica é a semântica do sistema, até que eventualmente alcançamos sistemas para os quais uma única interpretação semântica é praticamente (mas nunca em princípio) ditada.” (Trad. minha) (DENNETT, 1987, pp. 30-1). Neste sentido, a sofisticação semântica de um sistema está estreitamente ligada à plasticidade comportamental de um sistema dentro de um nicho ecológico, visto

O Princípio da Acumulação do Design parece trazer à tona implicações que vão contra a conclusão de Searle segundo a qual um organismo ou estrutura necessite reproduzir os aspectos causais do cérebro para se que possa ter intencionalidade, uma vez que o projeto do robô imaginado acima não assume que determinadas regularidades causais devam ser satisfeitas em sua estrutura final.

Poder-se-ia afirmar, no entanto, que o robô possui apenas intencionalidade derivada de seu criador, de modo que a descrição no nível semântico de seu comportamento deve ser feita em um contexto metafórico, isto é, o robô deve ser tratado *como se* tivesse intencionalidade. O robô, com efeito, não teria intencionalidade intrínseca, de modo que nenhum estado semântico poderia ser atribuído a ele em um contexto real¹².

Assumir que o nosso robô tenha intencionalidade derivada de nós parece, no entanto, indicar algumas dificuldades no que diz respeito à plasticidade comportamental visada pelo projeto inicial, visto que qualquer descrição semântica metafórica do sistema teria que ser derivada da nossa própria intencionalidade intrínseca. No entanto, conforme aponta Dennett (1987, p. 297), o grau de sofisticação do nosso robô exige que ele seja capaz de direcionar seu comportamento de modo que possa lidar efetivamente com adversidades ambientais, e estas adversidades, por sua vez, não poderiam ser previstas em toda sua amplitude por nós, criadores do robô. Neste sentido, a atribuição metafórica de intencionalidade ao robô parece apresentar dificuldades, uma vez que este parece ser capaz de produzir determinados comportamentos direcionados ao mundo, comportamentos estes que não obtêm sua direcionalidade na intencionalidade de seu criador, mas sim na intencionalidade derivada do

que, ainda que a intencionalidade deste sistema não seja intrínseca, a pressuposição de estados semânticos torna-se útil para a predição de seu comportamento. O exemplo do termostato pode ser utilizado em analogia com o caso do robô, isto é, na medida em que nós, os criadores do robô, adicionamos estruturas formais que aumentam sua sofisticação semântica, mais nos aproximaremos do resultado desejado pelo projeto inicial.

¹² Utilizo o termo “contexto real” em oposição a “contexto metafórico”.

próprio robô.

Tendo em vista a definição de intencionalidade apresentada no início deste trabalho, parece ser possível dizer que o robô, ainda que possua intencionalidade derivada, a saber, derivada das intenções que constituem o projeto de seu criador, pode ser capaz de empreender comportamentos que apresentam intencionalidade, mais especificamente, direcionalidade ao mundo, sem que de fato sua intencionalidade (ou direcionalidade) tenha que ser intrínseca no sentido em que Searle a define. A descrição intencional (e semântica) de um sistema, portanto, parece não requerer a existência de algo como a intencionalidade intrínseca¹³, o que parece confrontar a afirmação de Searle de que o robô enquanto mera reprodução sintática da operação do cérebro humano só teria estados intencionais em um sentido metafórico.

Um outro problema a se considerar com a tese distintiva da intencionalidade diz respeito a pressuposição de Searle de que a intencionalidade só pode ser intrínseca em um sistema uma vez que certas capacidades causais sejam satisfeitas. Esta afirmação parece estar comprometida com uma forma de *essencialismo* em relação às características biológicas de um sistema. Uma posição essencialista pressupõe que para que uma determinada característica biológica X seja de fato uma característica de um organismo, é preciso que um conjunto limitado de propriedades esteja presente neste organismo de modo que possamos dizer efetivamente se este sistema possui ou não a característica X.

¹³ Dennett (1987) acredita que a objeção de Searle não está preocupada de fato com os aspectos semânticos dos estados mentais, mas sim com uma característica mais fundamental, a saber, os *qualia*. A descrição semântica de um estado mental, dentro da filosofia de Dennett, parece não ser dependente de nenhuma propriedade intrínseca do cérebro, mas sim de uma postura interpretativa adotada frente ao comportamento de um sistema. Neste sentido, parece não haver contradições entre a atribuição de estados mentais com conteúdo semântico a um sistema que reproduza o comportamento de um ser humano através da operação mecânica e sintática de símbolos. Searle teria, portanto, na visão de Dennett, confundido a questão da possibilidade de atribuição de estados semânticos (e, portanto, intencionais) a um sistema que realize operações sintáticas com a possibilidade destes últimos darem origem aos aspectos qualitativos dos estados mentais.

Dentro do contexto acima, a intencionalidade abordada a partir de uma pressuposição essencialista assumiria que houve um momento *determinado* na história evolutiva no qual um sistema deixou de ser um sistema não-intencional para se tornar um sistema intencional. Este tipo de posicionamento, todavia, enfrenta dificuldades semelhantes a uma classificação essencialista das espécies, isto é, como podemos determinar um fator exato que nos permita distinguir uma espécie de outra? Em outras palavras, quais são os fatores que nos permitem apontar para determinado momento na história evolutiva e afirmar que neste mesmo momento houve a separação efetiva entre os grandes símios e o gênero *Homo*?¹⁴. Uma vez que se assuma uma postura essencialista, uma investigação sobre o surgimento da intencionalidade e a delimitação de sistemas intencionais e sistemas não-intencionais parece estar sujeita a estes mesmos questionamentos. Neste sentido, se trasladarmos estes questionamentos para o caso do robô, notar-se-ia que não existe um momento exato na construção do robô que nos permita dizer que este tenha se tornado um sistema capaz de produzir estados intencionais efetivos, já que o Princípio da Acumulação do Design pressupõe um desenvolvimento gradativo do sistema, de modo que, dentro da abordagem de Searle, a delimitação dos aspectos causais capazes de produzir intencionalidade tornar-se-ia uma tarefa extremamente difícil.

FUNÇÃO E SIGNIFICADO

A presente análise demonstrou até aqui que a tese de Searle, na qual a distinção entre intencionalidade intrínseca e derivada é defendida, parece não ser sustentável uma vez que assumamos como ponto de partida um estudo baseado nos pressupostos de uma proposta adaptacionista da

¹⁴ “Espécies “bem definidas” certamente existem – o propósito do livro de Darwin é explicar as suas origens – mas ele nos desencoraja de tentar achar uma definição baseada em “princípios” [*principled definition*] do conceito de espécie.” (Trad. minha) (DENNETT, 1995, p. 45). Assim, como aponta Dennett, a proposta de Darwin não nos impede de ainda falarmos em espécies, mas impossibilita, por outro lado, definições exatas destas, ou seja, não existe um conjunto definido de propriedades que determine a essência de uma espécie. Neste sentido, o essencialismo seria incompatível com as explicações que se baseiam no princípio da seleção natural.

teoria da evolução. De forma resumida, um sistema capaz de produzir comportamentos direcionados ao mundo pode adquirir esta capacidade a partir de uma acumulação ordenada de competências regidas por processos mecânicos que, em uma análise mais elementar, realizam operações sintáticas como a comutação de 0's e 1's seguindo um conjunto de regras pré-estabelecidas. A conceptibilidade de um sistema nos moldes do robô descrito acima demonstra que não é possível “desenhar” uma linha capaz de distinguir entre o que se pode entender por intencionalidade intrínseca e o que se pode entender por intencionalidade derivada, visto que, no caso do robô, ainda que este tenha sua intencionalidade derivada de nós, agentes intencionais (no sentido intrínseco), sua intencionalidade derivada lhe conferiria a capacidade de produzir estados intencionais que não obtêm seu conteúdo representativo (ou semântico) na nossa intencionalidade, já que a adversidade ambiental do nicho ecológico no qual o robô está inserido exigiria que previssemos um número infinito de situações, projeto este que não seria realizável em uma perspectiva prática¹⁵.

Torna-se claro, neste ponto, que a reprodução sintática do sistema cognitivo de um ser humano é capaz de engendrar um sistema capaz de comportar-se como se tivesse estados intencionais com conteúdo semântico, ou, simplesmente estados intencionais que possuem significado. Neste sentido, sistemas que possuem intencionalidade derivada podem produzir estados intencionais que têm seu conteúdo derivado da sua própria intencionalidade derivada, o que parece excluir a hipótese segundo a qual o significado seria atribuído ou criado por um sistema com intencionalidade intrínseca. Se a atribuição de significado, neste contexto, não é uma propriedade intrínseca do cérebro humano, de onde provém então o significado que atribuímos aos nossos próprios estados mentais e aos de outros indivíduos?

Este questionamento fica claro no exemplo da “Terra Gêmea” apresentado por Hilary Putnam (1975). Putnam pede-nos que imaginemos

¹⁵ É importante ressaltar que o projeto não seria realizável em uma perspectiva prática caso a distinção entre intencionalidade intrínseca e derivada fosse verdadeira.

o caso de um homem que tenha sido teletransportado para uma Terra Gêmea, isto é, um planeta fisicamente idêntico a Terra, mas que difere somente no fato em que a água dos rios e dos lagos é composta por moléculas de XYZ ao invés de moléculas de H₂O, embora o termo que os nativos da Terra Gêmea utilizam para se referir a XYZ seja o mesmo que nós utilizamos para nos referirmos a H₂O, isto é, “água”. Neste sentido, o indivíduo teletransportado da Terra para a Terra Gêmea, ao expressar seu desejo por um copo de água, estaria ele se referindo ao composto de H₂O ou ao composto de XYZ? O seu desejo seria um desejo por H₂O ou por XYZ?

Tomando o contexto aduzido por Putnam, poder-se-ia reformular o questionamento colocado acima em uma forma mais específica, enunciando-o nas seguintes palavras: o que determina se a minha crença de que o líquido que escorre pela torneira quando eu a abro é um composto de moléculas de H₂O e não de moléculas de XYZ? Isto é, como determinamos que o significado de minha crença diz respeito ao composto de H₂O e não ao composto de XYZ?

Dennett, em *The Intentional Stance*, argumenta em favor de uma abordagem teleológica da atribuição de significado como resposta ao questionamento proposto acima¹⁶. Segundo Dennett, não existe um fator determinante no âmbito do indivíduo que explicita se a minha crença expressa pela proposição “O líquido que escorre da torneira é água” está se referindo ao líquido cuja composição molecular é H₂O ou XYZ. Não há nenhum fato resultante da análise neurofisiológica do meu cérebro que me permita afirmar se o conteúdo de minha crença se refere à água₁ (H₂O) ou água₂ (XYZ)¹⁷. Isto ocorre porque o indivíduo enquanto

¹⁶ O conceito de teleologia pode ser compreendido aqui como a atribuição de *função* a um estado mental considerando a sua relação com o ambiente no qual está inserido. Assim, como se notará nas linhas que se seguem, a determinação do conteúdo mental da crença de que o líquido é H₂O ou XYZ se dará somente na medida em que consideramos a *finalidade* ou *função* que esta crença exerce nas disposições comportamentais deste sistema dentro do ambiente no qual está inserido.

¹⁷ É importante lembrar que Dennett é antes um *eliminativista* do que um *reducionista*

considerado de forma alheia ao seu ambiente não possui estados mentais com significado; a atribuição destes, na abordagem de Dennett, se dá apenas na medida em que consideramos a inserção deste indivíduo em um determinado ambiente, de modo que a minha crença de que o líquido que escorre da torneira é água₁ só pode ter este valor semântico consoante a consideração das disposições que possuo em relação a este ambiente e a *função* que o meu estado intencional exerce no direcionamento do meu comportamento em relação às adversidades do ambiente.

Neste contexto, a minha crença sobre o líquido que escorre da torneira quando a abro possuirá diferentes significados quando me encontro na Terra e quando me encontro na Terra Gêmea, visto que a atribuição de valor semântico a um estado intencional só se dá na medida em que considero a função que este estado exerce na economia comportamental do indivíduo em relação ao ambiente no qual ele está inserido. O significado, por conseguinte, é indeterminado no que se refere a uma análise limitada (*narrow*) do indivíduo enquanto tal; a sua atribuição só pode ser dada mediante a consideração da função que atribuímos a um organismo quando pretendemos interpretar seu comportamento.

Esta concepção teleológica da atribuição de significado corrobora com a análise de Dennett sobre o estatuto ontológico da intencionalidade humana na perspectiva de uma abordagem evolutiva, uma vez que a capacidade de atribuir significado aos nossos estados intencionais tem sua origem na intencionalidade do processo da seleção natural. Esta última, através do Princípio da Acumulação do Design, nos moldou de tal modo

no que se refere aos entes da *folk psychology* (crenças, desejos, etc.). Em poucas palavras, Dennett não acredita que haja uma correspondência estrita de uma crença ou de um desejo com um determinado padrão de impulsos eletroquímicos, por exemplo. Este tipo de posicionamento, no entanto, não o impede de assumir uma forma de realismo em relação à *folk psychology*. Esta discussão, no entanto, extrapola os objetivos deste artigo. Para uma exposição detalhada sobre o assunto, Cf. DENNETT, 1987, pp. 13-43. Ver também ZAWIDZKI, 2007, *The Intentional Stance*, pp. 31-54 e *Real Patterns*, pp. 143-52; ELTON, 2003, *Adopting a Stance*, pp. 31-67 e *Real Patterns*, pp. 68-98; e TEIXEIRA, 2008, pp. 29-66.

que podemos direcionar nosso comportamento em *função* das adversidades ambientais, o que nos confere a capacidade de representação do ambiente no qual estamos inseridos. Esta capacidade representativa nos fornece grande vantagem evolutiva na medida em que podemos engajar-nos no teste de hipóteses sobre como agir no ambiente sem que tenhamos que nos submeter a situações adversas¹⁸.

Tendo em vista as considerações acima, é possível notar que a distinção proposta por Searle parece apresentar dificuldades em casos como o do robô, no qual um sistema com intencionalidade derivada é capaz de produzir comportamentos que representam conteúdos semânticos sem que este conteúdo semântico tenha sua origem em outro sistema com intencionalidade intrínseca. Se Dennett estiver correto ao propor a experiência de pensamento do robô, poderíamos concluir que a intencionalidade humana, assim como a do robô, é uma intencionalidade derivada do processo seletivo que deu origem ao *Homo sapiens*¹⁹, visto que não é necessário que um sistema tenha algo como a intencionalidade intrínseca para que tenha comportamentos direcionados ao mundo²⁰. Uma

¹⁸ Como aponta Dennett (1996), esta vantagem evolutiva é retratada de forma precisa por uma frase enunciada por Karl Popper: “Karl Popper, de forma elegante, disse certa vez que este aprimoramento de design [a capacidade representativa] ‘permite que nossas hipóteses morram em nosso lugar’” (Trad. minha) (DENNETT, 1996, p. 88)

¹⁹ É importante notar que Dennett (1987) considera o processo de seleção natural como a única fonte de intencionalidade intrínseca apenas no sentido em que sua intencionalidade não tem origem em outro lugar. A seleção natural, entretanto, não possui objetivos ou direcionamentos ao longo de seu processo tal como seres humanos ou animais exibem em seu comportamento. Conforme notado anteriormente, o processo da seleção natural, na perspectiva de Dennett, é um processo algorítmico, e, portanto, mecânico, simples e sem inteligência. Em outras palavras, a nossa capacidade de direcionar nosso comportamento ao mundo é uma capacidade derivada de um longo processo de seleção mecânica e sem inteligência, processo este que por si só não tem nenhuma capacidade de direcionamento tal como seres humanos ou animais.

²⁰ A intencionalidade do homem é, por um lado, intencionalidade derivada e, por outro lado, intencionalidade intrínseca. Isto fica evidente na medida em que consideramos que a intencionalidade apresentada pelo homem é uma *intencionalidade derivada* do processo de seleção natural e que ao mesmo tempo é uma intencionalidade capaz de atribuir ou “criar” significados, o que a caracterizaria, em contraposição à primeira consideração, como *intencionalidade intrínseca*. Neste sentido, a distinção entre intencionalidade intrínseca

abordagem teleológica do significado dos estados intencionais parece, neste contexto, conciliar o fenômeno da intencionalidade com uma explicação naturalista da mente que segue os pressupostos da teoria da evolução e do funcionalismo, sem que tenhamos que recorrer às distinções propostas por Searle entre intencionalidade intrínseca e intencionalidade derivada. Neste sentido, os desenvolvimentos do funcionalismo rotulados por Searle como Inteligência Artificial Forte não parecem estar comprometidos com uma abordagem na qual a atribuição de conteúdo semântico a sistemas formalmente ou sintaticamente construídos careça de explicação.

Abstract: The problem of intentionality in philosophy of mind regards the way human and other animals minds are related to the world, a problem that requires a careful analysis by a philosophical approach whose objective is to explain this topic. In *Mind, Brains and Programs*, John Searle offers a scathing criticism to functionalists approaches of intentionality and mental states' meaning which rely on what he calls "Strong Artificial Intelligence" program research. On Searle's view, the mere syntactic reproduction of the logical relationships of the human brain is not sufficient to reproduce a system with intentionality and meaning. Opposing Searle, Daniel Dennett argues for the thesis that intelligent apparently designed systems may have its origin in simple and mechanical process such as algorithmic process, which seems to indicate at first the possibility of an explanation of intentionality following these presuppositions. Having in mind Dennett's proposal, the objective of this article is to expose some aspects of Dennett's theory of intentionality and to point out some considerations that may enable a conciliation of a functionalist theory of mind and the phenomenon of intentionality, adducing for the process of natural selection as the fulcrum for this conciliation.

Keywords: evolution; functionalism; intentionality; meaning; Dennett.

e intencionalidade derivada apresentada por Searle não parece ser uma distinção adequada mediante as considerações sobre a teoria da evolução feitas até aqui.

Referências

- DENNETT, D. *The Intentional Stance*. Cambridge: MIT Press, 1987.
- _____. *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. New York: Simon & Schuster, 1995.
- _____. *Kinds of Minds: Towards an Understanding of Consciousness*. New York: Basic Books, 1996.
- ELTON, M. *Daniel Dennett*. Cambridge: Polity Press and Blackwell Publishing, 2003.
- PUTNAM, H. (1973). *The Nature of Mental States*. In: CHALMERS, D. (Org.). *Philosophy of Mind: Classical and Contemporary Readings*. New York: Oxford University Press, 2002.
- PUTNAM, H. (1975). *Mind, Language and Reality*. New York: Cambridge University Press, 1997.
- SEARLE, J. (1980). *Minds, Brains and Programs*. In: HOFSTADTER, D.; DENNETT, D. (Orgs.). *The Mind's I*. New York: Basic Books, 2000.
- _____. (1984). *Intencionalidade*. São Paulo: Martins Fontes, 1995.
- _____. (1992). *A Redescoberta da Mente*. São Paulo: Martins Fontes, 1997.
- TEIXEIRA, J. *A Mente Segundo Dennett*. São Paulo: Perspectiva, 2008.
- TURING, A. (1950). *Computing Machinery and Intelligence*. In: COPELAND, B. *The Essential Turing*. New York: Oxford University Press, 2004.
- ZAWIDZKI, T. *Dennett*. In: *Oneworld Thinkers*. Oxford: Oneworld Publications, 2007.