

O Experimento de Pensamento do Quarto Chinês: a Crítica de John Searle à Inteligência Artificial Forte

RESUMO

Será que um dia serão desenvolvidos computadores digitais capazes de pensar de modo *similar* ao nosso? Ou será que, independentemente da tecnologia, os computadores digitais estarão sempre limitados a manipular dados *sem* compreendê-los? Neste trabalho, apresentarei duas concepções antagônicas de Filosofia da Mente: a Inteligência Artificial Forte (IA Forte), que responde afirmativamente à primeira questão, e a crítica de John Searle a esta corrente, que, por sua vez, responde de maneira afirmativa à segunda questão. Para tanto, iniciarei o artigo apresentando o famoso *jogo da imitação* proposto por Alan Turing (1950) para decidir se uma máquina é ou não inteligente, jogo este que ficou conhecido como *teste de Turing*. Logo em seguida, analisarei minuciosamente o experimento mental do *quarto chinês* (proposto em 1980 por John Searle), que é uma crítica ao teste de *Turing* e ao programa de pesquisa da Inteligência Artificial Forte (IA Forte).

Palavras-chave: Teste de turing; Inteligência artificial; Quarto chinês; John Searle.

ABSTRACT

Will one day developed digital computers be able to think in a *similar* way we human beings do? Or, being independent of technology, will they always be limited to manipulate data without comprehend them? In this work, I will present two opposite conceptions of Philosophy of Mind: the Strong Artificial Intelligence, which responds positively to the first question, as well as John Searle's critique to this line, which responds positively to the second question. Therefore I will begin the article presenting the famous *imitation game* proposed by Alan Turing (1950) in order to decide if a machine is intelligent or not. This game is known by *Turing's test*. Then I will exhaustively analyze the mental experiment of the *chinese room*, proposed by John Searle in 1980, which is a strong critique to the *Turing's Test* and to the research program of the Strong Artificial Intelligence.

Key words: Turing's test; Strong artificial intelligence; Chinese room; John Searle.

* Mestrando em Filosofia pela Universidade Federal do Ceará (UFC / Capes). Agradeço ao Professor Dr. Tárík Prata, da Universidade Federal de Sergipe (UFS), pelas críticas e sugestões a este trabalho.

Um dos principais temas da ficção científica do século XX, seja na literatura ou no cinema, é o que aborda a Inteligência Artificial (IA) e suas conseqüências, positivas e negativas, para a humanidade. O maior temor em relação às conseqüências negativas diz respeito a um tipo de “revolta” das máquinas contra os humanos: em algum momento de um futuro próximo, as máquinas – tanto na forma de computadores como de robôs – tornar-se-iam *realmente* inteligentes ou, de maneira mais correta, possuiriam várias características que até então eram tidas por muitos como exclusivas de nossa espécie. Com isso, em algumas histórias do gênero, as máquinas se *sentem* exploradas e acabam se *rebelando* contra seus criadores. Foi pensando nisso que um dos maiores autores de ficção científica de todos os tempos, Isaac Asimov, criou as Três Leis da Robótica:¹

- 1) Um robô não pode prejudicar um ser humano ou, por omissão, permitir que o ser humano sofra dano;
- 2) Um robô tem de obedecer às ordens recebidas dos seres humanos, a menos que contradigam a Primeira Lei;
- 3) Um robô tem de proteger sua própria existência, desde que essa proteção não entre em conflito com a Primeira e Segunda Leis.

Após o que foi dito, resta uma importante pergunta de cunho teórico-prático: algum dia a IA da ficção científica seguirá o caminho de outras grandes tecnologias de nossa civilização, como o submarino, isto é, deixará de ser *sonhos* para se tornar realidade? Marvin Minsky, um dos expoentes da IA, acha que é só questão de tempo para as máquinas não só atingirem as

capacidades humanas, mas mesmo ultrapassá-las, chegando ao ponto de afirmar que um dia aquelas terão orgulho de nós, seus criadores.² Para Minsky, “cérebros são computadores formados de carne”. Já Searle, que é um dos maiores críticos desse modelo de IA,³ argumenta que quando um computador realiza uma tarefa que também é realizada por um humano, ele a realiza de maneira totalmente diferente, já que aquele não possui *consciência* nem *compreensão* acerca do que está executando.⁴ Em contraposição a Searle, está a corrente fisicalista que mais se desenvolve nos últimos tempos, o Funcionalismo, que, de maneira geral, é famoso por sua analogia computador-homem: o *software* está para o *hardware* assim como a mente está para o cérebro. Segundo essa visão, não importa de qual material seja constituído o hardware, mas sim a sua *organização funcional*. Isso significa que nossos cérebros estão organizados de maneira tal que surgem estados mentais a partir de seu funcionamento, mas *não* existe nenhum impedimento de que outra constituição física, como circuitos de silício, por exemplo, esteja organizada de modo semelhante e produza uma mente. Sendo assim, o que importa para que um estado seja funcional é sua *relação* entre determinados dados de entrada (*inputs*) e suas respectivas respostas (*outputs*). Dessa maneira, diferentemente da Teoria de Identidade, que considera estados mentais como *idênticos* a estados cerebrais, o funcionalista considera que um estado mental é funcional, isto é, idêntico a um papel causal, que, por sua vez, influenciará determinados comportamentos e estados mentais posteriores.

Conclui-se, segundo os adeptos do funcionalismo, que não há motivos para duvidar que um computador possa ter uma mente, basta

¹ As Três Leis da Robótica foram mencionadas pela primeira vez no livro *Eu, Robô*, de 1950. Em um livro posterior, *Os robôs do amanhecer*, Asimov lança outra Lei, que é a mais básica de todas, a saber, a Lei Zero: um robô não pode fazer mal à humanidade e nem, por omissão, permitir que ela sofra algum mal.

² Na verdade, em outras declarações Minsky chega a cogitar um futuro bem menos digno para a nossa espécie, já que talvez as máquinas não nos queiram nem mesmo como animais de estimação! É da seguinte maneira que Guitta Pessis-Pasternak apresenta o entusiasta da IA: “Marvin Minsky, famoso informaticista do MIT (Massachusetts Institute of Technology), e um dos fundadores da área de Inteligência Artificial americana, prevê para as décadas que virão o desenvolvimento de máquinas dotadas de “inteligência geral de um ser humano médio”. E, aos poucos, essas obras-primas “assumirão a sua própria educação. Em alguns anos, elas alcançarão o nível do gênio, e depois o seu poder será incalculável”; a tal ponto que *teremos sorte se elas resolverem nos conservar como animais domésticos*. (PESSIS-PASTERNAK, 1992, p. 207. Grifos meus).

³ Existem diferenças entre Inteligência Artificial Fraca e Inteligência Artificial Forte. A IA fraca parte da tese que os computadores podem nos ajudar na compreensão do ser humano, como, por exemplo, no que diz respeito à cognição humana. Mas trata-se apenas de simulação, e não de *realização* propriamente dita. Já para a IA forte, simular uma mente é o *mesmo* que ter uma mente. Portanto, qualquer máquina que passe no teste de Turing tem uma vida mental semelhante a nossa.

para isso que ele possua os mesmos estados funcionais de um cérebro humano.

Em 1950, Alan Turing, que é considerado, juntamente com John Von Neumann, o pai do computador moderno,⁵ publicou um artigo em que tentava responder a questão “podem as máquinas pensar?”.⁶ Para responder tal pergunta, Turing propôs um teste, que será analisado em detalhes na seção seguinte.

O Teste de Turing

Turing inicia seu artigo sobre Computação e Inteligência propondo o *jogo da imitação* como substituto para a pergunta “podem as máquinas pensar?” (TURING, 1996, p. 21). Um homem (A), uma mulher (B) e um interrogador ou interrogadora (C) participam do jogo da imitação. O interrogador está separado dos outros dois jogadores. Cada pessoa tem um objetivo diferente no jogo: C faz perguntas aos outros dois participantes e seu objetivo é identificar corretamente qual o homem e qual a mulher; o objetivo de A, por sua vez, é induzir C ao erro; por fim, a meta de B é dar respostas que auxiliem C na correta identificação. Portanto, A poderia fornecer respostas⁷ típicas femininas, como dizer, por exemplo, que tem cabelo comprido e usa vestido. Por sua vez, B poderia dizer para não acreditar em A, e que ela era a verdadeira mulher. Caso tenha sucesso, C acabará o jogo com o seguinte veredicto: A é homem e B é mulher, situação esta em que B também atinge seu objetivo. Por outro lado, A sairá vitorioso perante seus companheiros de jogo se C disser que A é mulher e B é homem.

Após propor o jogo da imitação, Turing pergunta “o que acontecerá quando uma máquina ocupar o lugar de A nesse jogo?”, ou seja, o que ocorrerá se um computador tentar se passar por ser humano (TURING, 1996, p. 22). Para Turing, responder essa pergunta corresponde a responder se as máquinas podem pensar:

[...] isso significa que um programa de computador é inteligente caso não fosse possível ao interrogador diferir, de maneira estatisticamente significativa, suas respostas das respostas fornecidas pelo jogador humano.⁸

Isso ficou conhecido como teste de Turing, e até hoje é oferecido um prêmio⁹ para quem desenvolva um programa de computador que passe no teste.

Turing estava ciente do conhecimento precário sobre as bases neurobiológicas da consciência, mas, mesmo assim, pensava que era possível chegar a uma conclusão em relação ao teste de Turing sem o auxílio das futuras descobertas neurocientíficas sobre a consciência:

Não quero dar a impressão de que penso não existir nenhum mistério no que diz respeito à consciência. Existe, por exemplo, algo assim como que um paradoxo vinculado às tentativas de localizá-la. Mas não acredito que tais mistérios tenham de ser necessariamente resolvidos antes de podermos responder à pergunta que nos preocupa neste artigo. (TURING, 1996, p. 41).

⁴ Como será visto adiante com o experimento mental do quarto chinês.

⁵ Na verdade, o projeto do computador digital é antigo, tendo sido idealizado entre 1828 e 1839 por Charles Babbage. Apesar das idéias essenciais intuídas por Babbage, o precursor do computador digital, o qual ele chamou de *máquina analítica*, não chegou a ser construído por causa da insuficiente tecnologia da época (TURING, 1996, p.29).

⁶ O artigo chama-se *Computing Machinery and Intelligence* e foi publicado em 1950, no prestigioso periódico britânico **Mind**. Este artigo foi traduzido para o português por Fábio de Carvalho Hansen com o título de *Computação e Inteligência*, In: J. F. Teixeira, *Cérebros, máquinas e consciência: uma introdução à filosofia da mente*. (São Carlos: Editora da UFSCar, 1996), p. 19-60.

⁷ Para que haja o mínimo de influência, as respostas fornecidas a C devem ser digitadas (ou datilografadas, como melhor convinha à tecnologia utilizada na época, a máquina de escrever).

⁸ Vale a pena ressaltar que Turing não estava afirmando que *todos* os computadores digitais seriam capazes de passar no teste. Além disso, ele não disse que os computadores de sua época, isto é, da segunda metade do século XX, teriam bom desempenho no jogo, mas, sim, que poderíamos *conceber* computadores com essa capacidade. (TURING, 1996, p. 25-26).

⁹ Ele é chamado de prêmio Loebner e foi proposto pelo inventor e industrial Hugh Loebner, o qual ofereceu “um prêmio de 100 mil dólares para o primeiro programa de computador que passe na sua versão do teste de Turing. Embora *nenhum* programa tenha ganhado o Grande Prêmio, há um prêmio menor de 2 mil dólares para o programa de computador *mais humano* da competição, e isso atrai boas tentativas.” (WHITBY, 2004, p. 32. Grifos meus).

O Debate Acerca do Quarto Chinês

Trinta anos após o artigo *Computação e Inteligência*, Searle publica um artigo no qual propõe seu famoso experimento de pensamento do quarto chinês, que é uma crítica à *Inteligência Artificial Forte* (IA Forte).¹⁰ Segundo Moural (2003), é possível fazer a distinção entre a) o experimento de pensamento propriamente dito, b) o argumento teórico relacionado a este e, por último, c) algumas questões de cunho periférico.

a) O experimento de pensamento do quarto chinês

A idéia do quarto chinês surgiu por causa de um *software*, especializado em responder perguntas relacionadas a uma história, desenvolvido por Schank e Abelson. Apesar desse *software*, em particular, ter sido o ponto de partida de Searle, o experimento do quarto chinês não se limita ao mesmo: o quarto chinês se aplica igualmente bem a qualquer máquina de Turing, portanto a qualquer *software* que tenha a pretensão de possuir fenômenos mentais similares aos humanos. O quarto chinês tem o intuito de atacar duas teses centrais da IA, a saber:

- 1) Que é lícito afirmar categoricamente que a máquina realmente *compreende*¹¹ a história e é capaz de prover respostas adequadas às perguntas que lhe são feitas;
- 2) Que a máquina e seu programa são capazes de *explicar* a capacidade humana para se entender uma história e responder às perguntas sobre a mesma.

O experimento de pensamento é o seguinte. Searle é um falante de inglês que não compreende nada de chinês¹² e está trancado dentro de um quarto. Searle recebe algumas tiras de papel escritas em chinês por meio de

uma abertura. Apesar de não compreender o que está escrito, Searle tem acesso a um livro que contém um conjunto de regras, escrito em inglês, que correlaciona a tira de papel que ele recebeu a um outro conjunto de tiras de papel, também escritas em chinês, que está à sua disposição no interior do quarto. As instruções do livro são do tipo “ao receber tira contendo x, coloque pela abertura da porta a tira contendo y”. Vale a pena ressaltar que tanto x como y são mensagens com caracteres em chinês, e que Searle só entende a instrução, escrita em inglês, que correlaciona x com y. A correlação entre o símbolo recebido e o símbolo enviado é puramente *formal*, isto é, Searle os identifica unicamente pelos seus formatos. Sem que Searle *saiba*, ele está respondendo corretamente, do ponto de vista de um observador falante de chinês que está do lado de fora do quarto, às perguntas que lhe foram fornecidas. Portanto, o quarto chinês se comporta de maneira semelhante a um computador: a tira que Searle recebe corresponde ao dado de entrada ou *input*; o livro de instruções é análogo ao programa ou *software* de computador; o próprio Searle faz o papel da unidade central de processamento (CPU); e, por fim, a tira de papel enviada por Searle corresponde ao dado de saída ou *output*. O ponto central do experimento de pensamento é que Searle, imitando o funcionamento de um computador, pode responder corretamente às perguntas em chinês *sem compreender* nada de chinês. Sendo assim, Searle ataca a tese 1 da IA mostrando que um programa de computador não possui compreensão genuína, já que assim como Searle não entende chinês, o computador também não entende a história sobre a qual ele responde, por mais que essas respostas sejam corretas e semelhantes às que seriam dadas por um ser humano. Portanto, manipular símbolos não é o mesmo que compreender símbolos, isto é, a manipulação de uma sintaxe é condição necessária, mas não suficiente para a existência de uma semântica. Podemos ver que o quarto

¹⁰ *Minds, brains, and programs*, publicado em **The Behavioral and Brain Sciences**, Volume 3, p. 427-424 (1980) Este artigo foi traduzido para o português por Cléa Regina de Oliveira Ribeiro com o título de *Mentes, Cérebros e Programas*, In: J. F. Teixeira, *Cérebros, Máquinas e Consciência: Uma introdução à Filosofia da Mente* (São Carlos: Editora da UFSCar, 1996), pp. 61-93.

¹¹ Compreensão nesse sentido não significa que a máquina é apenas capaz de *imitar* o comportamento humano, mas que, de fato, a máquina *possui* estados mentais semelhantes aos nossos.

¹² O que chamamos de língua chinesa é na verdade formada por uma família de línguas pertencentes ao ramo sino-tibetano. O mandarim e o cantonês são dois dos dialetos mais conhecidos da língua chinesa.

chinês passa no teste de Turing, mesmo sem compreender uma única palavra em chinês.

Em relação à tese 2, Searle afirma que compreender o funcionamento do *software* não seria de grande contribuição para explicar o nosso próprio entendimento de histórias, pois apesar do programa responder corretamente às perguntas, ele ainda assim não compreende nada da história em questão.

Para fortalecer ainda mais sua crítica, Searle pede que modifiquemos um pouco o experimento acima: as tiras de papel, tanto das perguntas como das respostas, estarão grafadas em inglês. A diferença crucial deste segundo cenário em relação a o primeiro é que enquanto neste havia *apenas* a manipulação formal de símbolos, naquele há, além da mera manipulação simbólica, *compreensão* dos símbolos manipulados. Dito de outra forma, utilizando o jargão da Lingüística: no primeiro cenário há somente sintaxe, diferentemente do segundo, no qual está presente tanto a sintaxe como a semântica. Conclui-se, então, que apesar de um computador poder *imitar* o nosso comportamento,¹³ ele *não funciona* como nossa mente – aliás, ele nem possui uma mente –, já que apenas manipula formalmente os símbolos sem ter qualquer conhecimento sobre seu significado.

b) O argumento relacionado ao quarto chinês

Searle pretende demonstrar, através de seu argumento, que a implementação do programa correto é *insuficiente*, por si só, para gerar o entendimento, já que não há intencionalidade na pura manipulação formal de símbolos. O argumento de Searle seria composto de quatro premissas e quatro conclusões¹⁴ derivadas destas. As premissas são:

P1) O cérebro causa a mente (tese da *redução causal* dos estados mentais a estados cerebrais, a qual é amplamente sustentada pelas evidências provenientes das neurociências);

P2) A sintaxe não é suficiente para a semântica (essa premissa constitui uma verdade conceitual);

P3) Programas de computador são totalmente definidos por sua estrutura formal ou sintática (essa premissa é verdadeira por definição);

P4) A mente possui conteúdos semânticos (segundo Searle, isso é um fato óbvio). Quatro conclusões são geradas a partir das premissas acima, quais sejam:

C1) Nenhum programa é, por si mesmo, suficiente para produzir uma mente. Isto é, além de programas não serem mentes, eles não são suficientes, por si próprios, para gerarem uma mente (derivada de P2, P3 e P4);

C2) O modo como o funcionamento cerebral causa a mente não pode ser unicamente em decorrência da implementação de um programa de computador (derivada de P1 e C1);

C3) Qualquer coisa que cause uma mente tem que ter capacidades causais pelo menos equivalentes àquelas do cérebro (derivada diretamente de P1);

C4) A implementação de um programa de computador não seria capaz de dotar, por si mesma, uma máquina de estados mentais semelhantes aos nossos, já que esta teria que ter capacidades causais equivalentes às capacidades do nosso cérebro (derivada de C1 e C3).

Com o passar do tempo, Searle reformula, de maneira cada vez mais concisa, seu argumento. Searle apresenta em seu livro *Mistério da Consciência* (1998, pp. 38-39) o núcleo central do argumento, o qual consiste apenas das premissas P2, P3, P4 e da conclusão C1:

1. Programas são totalmente sintáticos (P3);
2. As mentes têm uma capacidade semântica (P4);
3. Sintaxe não é a mesma coisa que a semântica, nem é, por si só, suficiente para garantir um conteúdo semântico (P2). Consequentemente, programas não são mentes (C1).

¹³ Como já foi dito, uma máquina que se comporte, por meio das respostas fornecidas a um interrogador, de maneira indistinta em relação a um ser humano passa no teste de Turing. Isso significa, segundo os defensores da IA Forte, que tal máquina teria uma mente com conteúdos mentais semelhantes aos nossos, isto é, ela seria capaz, entre outras coisas, de compreender verdadeiramente uma história.

¹⁴ Conferir Moural (2003), p. 219-220.

c) Questões Periféricas

Para a IA Forte a implementação de um programa correto¹⁵ em *qualquer tipo de hardware*¹⁶ bastaria para produzir uma mente. Searle afirma que o slogan da IA Forte é o seguinte: a mente está para o cérebro assim como o *software* está para o *hardware*. Searle pretende mostrar que o programa de pesquisa da IA Forte está fadado ao fracasso, porque esse programa leva a conclusões absurdas.

Searle se utiliza do conceito de intencionalidade para criticar a tese de que a simples implementação de um programa correto seria suficiente para causar uma mente, já que sua teoria do significado é fundamentada na intencionalidade, mais precisamente, na sua concepção da intencionalidade intrínseca dos estados mentais. A maioria¹⁷ dos estados mentais é intencional. Estados intencionais, por sua vez, não são definidos por sua forma – como o são os programas –, mas pelos seus conteúdos. Já a manipulação formal de símbolos que é característica dos programas de computador não possui qualquer tipo de intencionalidade, sendo, portanto, totalmente desprovida de significado. Como Searle assinala em diversas passagens de seus trabalhos:¹⁸ os programas têm apenas sintaxe, mas não semântica.

Apesar da diversidade de propostas, pode-se afirmar que a maioria dos cientistas e filósofos atuais defende *alguma* versão fisicalista¹⁹ em se tratando da relação existente entre mente e sistema nervoso, particularmente, entre mente e cérebro. São muitas as evidências provenientes das neurociências que

apóiam a premissa I (P1) do argumento de Searle, ou seja, que a mente é *causada* pelo cérebro. Searle se utiliza dessa premissa para apontar mais uma falha na analogia da mente como um programa, já que esta é *produzida* pelo cérebro, enquanto aquele não é um produto do *hardware*.

Searle acusa os partidários da IA forte de serem, apesar de soar de maneira contraditória, behavioristas e dualistas, além de confundirem simulação com duplicação. Ao behaviorismo se deve à tese de que é possível descrever as propriedades mentais de um sistema unicamente em função da observação de seu comportamento. Isso é bem característico do teste de Turing, pois o que está sendo avaliado são as respostas (*outputs*) que a máquina fornece para as perguntas (*inputs*) feitas pelo interrogador: afirmar que uma máquina se *comporta* como um humano é o mesmo que dizer que ela possui estados mentais semelhantes aos dos homens. A perspectiva do interrogador em relação à máquina é uma perspectiva de terceira pessoa e, a partir desta, ele julgará se a máquina possui estados mentais. Contudo, Searle afirma que o “comportar-se como humano” não é garantia para ter uma “mente semelhante à humana.”²⁰

[...] ontologicamente falando, comportamento, papel funcional e relações causais são irrelevantes para a existência de fenômenos mentais conscientes²¹ (SEARLE, 1997, p. 103).

Já que Searle defende que os estados mentais têm uma ontologia de primeira pessoa

¹⁵ A expressão programa correto significa, neste contexto, um *software* que descreva formalmente os estados mentais através de um processamento de informação. Conferir Moyal (2003), p. 221.

¹⁶ Para a corrente funcionalista, o *software* tem *primazia* sobre o *hardware*, pois aquele independe da constituição física deste. Dito de outra maneira, a *organização* dos elementos de um sistema é mais importante do que a *constituição física* destes elementos. Daí se conclui que circuitos de silício, só para citar um exemplo, organizados de maneira formalmente semelhante aos neurônios de nosso cérebro serão responsáveis, causalmente falando, pela produção de uma mente similar à nossa. Tese esta, como já foi visto, que Searle combate veementemente, já que, segundo ele, a organização destes circuitos de silício não reproduziria as capacidades causais do cérebro. Ver o argumento acima exposto, notadamente as conclusões C3 e C4.

¹⁷ Nem todo estado mental é intencional: “Posso, por exemplo, simplesmente estar num estado de espírito de depressão ou exultação sem estar deprimido ou exultante por qualquer coisa em particular. Nesses casos, meu humor, como tal, não é intencional” (SEARLE, 1997, p. 188).

¹⁸ “No jargão linguístico, eles têm apenas sintaxe, mas não semântica” (SEARLE, 1996, p. 86).

¹⁹ Searle chega ao ponto de fazer a seguinte pergunta: “o que fez com que o “materialismo” parecesse ser o único enfoque racional da filosofia da mente?” (SEARLE, 1997, p. 10). Searle defende que sua teoria, o Naturalismo Biológico, não se enquadra nem como dualismo nem como fisicalismo, já que apesar das diferenças entre essas duas correntes elas incorrem no mesmo erro básico, a saber, se utilizam de uma terminologia obsoleta, do século XVII, que contrapõe o mental ao físico. Sendo assim, ele defende que tanto o dualismo quanto o fisicalismo se apóiam no que ele designa de *dualismo conceitual*. Searle afirma que “a tradição materialista é sólida, ubíqua e, ainda assim, evasiva”, sendo, erroneamente, chamada de ciência (SEARLE, 1997, p. 18).

²⁰ Isso está intimamente ligado à confusão entre *simulação* e *duplicação* que será apresentada logo a seguir.

²¹ Searle propõe, a partir do experimento mental do cérebro de silício (SEARLE, 1997, p. 98-105), o *princípio da independência de consciência e comportamento*, o qual afirma que o comportamento não é nem condição suficiente nem condição necessária para os fenômenos mentais.

e que o ponto de vista de primeira pessoa é, em termos epistêmicos, bem diferente do ponto de vista de terceira pessoa, pode-se perguntar qual o *fundamento empírico* que ele utiliza para afirmar que outros sistemas²² possuem estados conscientes. Sua resposta a esta questão é que os métodos empíricos utilizados por uma pessoa (ponto de vista de terceira pessoa) para inferir e estudar os fenômenos mentais (ontologia de primeira pessoa) de outro sistema consciente

[...] se fundamentam num princípio prático que empregamos alhures na ciência e no dia-a-dia: *mesmas causas-mesmos-efeitos*, e *causas semelhantes-efeitos-semelhantes*. (SEARLE, 1997, p. 112).

Portanto, para Searle, a descoberta de que outros organismos possuem estados mentais não se deve apenas à observação de seu comportamento, mas, sim, a uma *conexão* que existe entre o comportamento e a estrutura causal destes indivíduos. (SEARLE, 1997, p. 115): como os seres humanos compartilham a mesma fisiologia²³ (mesma causa), pode-se inferir, a partir do princípio prático supracitado, que eles possuem também experiências conscientes idênticas (mesmo efeito) às nossas. Devido a isso, Searle afirma que, do ponto de vista prático ou da vida real, não existe o que os filósofos da mente chamam do *problema de outras mentes*, pois

[...] não sustentamos uma "hipótese", "crença" ou "suposição" de que outras pessoas sejam conscientes. Ao contrário, temos determinadas maneiras de nos comportar advindas do *Background*, determinadas capacidades possibilitadas pelo *Background*, e essas são constitutivas de nossas relações com a consciência de outras pessoas. (SEARLE, 1997, p. 116).

De maneira semelhante, apesar de um cachorro, um chimpanzé ou um golfinho pertencerem a espécies diferentes da nossa, em muitos aspectos relevantes, para a questão tratada aqui, eles são fisiologicamente semelhantes ao *Homo sapiens*, daí a atribuição de fenômenos

mentais similares aos nossos a esses organismos: causas semelhantes-efeitos semelhantes. (PRATA, 2007, p. 184-185)

Outra crítica de Searle à IA Forte se deve a afirmação que, em princípio, qualquer substrato material devidamente organizado possui uma mente. Isso constitui uma posição dualista, pois as propriedades mentais estão dissociadas das propriedades físicas. Como já foi dito, Searle rejeita essa tese: já que o cérebro causa a mente (P1), as propriedades físicas do cérebro são relevantes para a produção das propriedades mentais. Isso significa que se algum substrato físico, que não seja o sistema nervoso, causa uma mente, ele tem que ter capacidades causais equivalentes às do cérebro (C3).

Por fim, Searle alega que há uma confusão entre *simulação* e *duplicação* que permeia o discurso dos defensores da IA Forte. Do ponto de vista de um observador externo, Searle se comporta, em seu experimento de pensamento, como um falante de chinês, apesar de não compreender uma única palavra neste idioma. Portanto, pode-se dizer que Searle simula a capacidade linguística de um falante de chinês, mas não a duplica, já que a capacidade linguística pressupõe a compreensão dos símbolos manipulados e esta não está presente²⁴. Essa distinção entre simular e duplicar fica melhor explicitada ao se constatar que um meteorologista não fica com os nervos à flor da pele ao empregar programas de computador para simular tempestades e terremotos. A calma do meteorologista é totalmente justificada em sua crença ou conhecimento de que simular tais intempéries da natureza *não implica* reproduzi-las. Da mesma maneira, Searle pergunta o porquê de alguém supor "que uma simulação computacional da compreensão de fato entenderia alguma coisa." (SEARLE, 1996, p. 88). Para simular a cognição, por exemplo, basta que o programa atue na transformação de um *input* em um *output* adequado, e isso é totalmente executável. Por conseguinte, um programa de computador se *limita* a simular fenômenos naturais e mentais, donde conclui-se que "confundir

²² Tanto pessoas quanto outros animais, como cachorros, chimpanzés e golfinhos.

²³ Desconsiderando-se, é claro, os casos patológicos e as pequenas diferenças individuais.

²⁴ "[...] a manipulação de símbolos formais por si só não tem intencionalidade: eles não têm significado, eles nem mesmo são manipulações de *símbolos*, uma vez que esses símbolos não significam nada". No jargão linguístico, eles têm apenas sintaxe, mas não semântica." (SEARLE, 1996, p. 86).

simulação com duplicação é o mesmo erro, seja com dor, amor, cognição, incêndio ou tempestade.” (SEARLE, 1996, p. 89).

Referências Bibliográficas

MOURAL, Josef. The Chinese Room Argument. In: SMITH, Barry (Ed.). *John Searle*. Cambridge University Press, 2003.

SEARLE, John. Minds, brains, and programs. *The Behavioral and Brain Sciences*, 3, 1980, p. 417-424.

SEARLE, John. Mentas, 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.

PRATA, Tárik. Subjetividade Ontológica na Filosofia da Mente de John Searle. *Philosophos*. v. 12, n. 2, 2007, p. 171-204, jul/dez.

SEARLE, John. *A Redescoberta da mente*. Tradução de Eduardo Pereira Ferreira. São Paulo: Martins Fontes, 1997.

_____. *O Mistério da consciência*. Tradução de André Yuji Pinheiro Uema e Vladimir Safatle. São Paulo: Paz e Terra, 1998.

PESSIS-PASTERNAK, Guitta. *Do caos à inteligência artificial*. Tradução de Luiz Paulo Rouanet. São Paulo: Editora Unesp, 1992.

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

TURING, Alan. Computing machinery and intelligence. Reimpresso In: HEIL, John. *Philosophy of mind: a guide and anthology*. Great Britain: Oxford University Press, 2004.

WHITBY, Blay. *Inteligência artificial: um guia para iniciantes*. Tradução de Cláudio Blanc. São Paulo: Madras, 2004.