

## UMA NOTA SOBRE O ARGUMENTO DO QUARTO CHINÊS DE JOHN SEARLE

João de Fernandes Teixeira<sup>1</sup>

ORCID iD: <https://orcid.org/0000-0002-4109-6139>

**Resumo:** O artigo propõe uma rediscussão do Argumento do Quarto Chinês formulado por John Searle em 1980. Após uma recapitulação dos delineamentos do argumento na primeira parte, uma discussão preliminar de uma objeção formulada por David Chalmers é apresentada. A terceira parte visa a apresentação de uma possibilidade relegada por Searle na apresentação de seu argumento: as máquinas e robôs semi-biológicos que estão em desenvolvimento na Universidade de Reading, Inglaterra. Esses dispositivos não foram considerados por Searle, que não mudou sua argumentação desde a formulação original na década de 1980, que se limitou a objetar ao modelo computacional de mente. Dessa forma argumentamos que o próprio naturalismo biológico de Searle constitui uma objeção ao seu argumento do quarto chinês.

**Palavras-chave:** Searle; Argumento do quarto chinês; Máquinas e robôs semi-biológicos; Modelo computacional da mente.

**Abstract:** The article proposes a re-discussion of the Chinese room argument formulated by John Searle in 1980. After a review of the outline of the argument, an objection formulated by David Chalmers is presented. The third part aims at presenting a possibility relegated by Searle in the formulation of his argument: the semi-biological robots and machines that are under development at the University of Reading, England. These devices were not considered by Searle who has not changed his argument since the original formulation in the 1980s, which was limited to objecting to the computational model of the mind. Thus, we argue that Searle's own biological naturalism constitutes an objection to his Chinese room argument.

**Keywords:** Searle; Chinese room argument; Semi-biological machines and robots; Computational model of the mind.

O argumento do quarto chinês, formulado por John Searle em 1980 foi por muito tempo uma poderosa máquina de guerra contra as pretensões da inteligência artificial de simular as atividades mentais humanas. Na época, as apostas dos pesquisadores se concentravam no modelo computacional da mente, ou seja, no que ficou conhecido posteriormente como G.O.F.A.I. (Good and Old Fashioned Artificial Intelligence).

Mas, será que passadas quase cinco décadas após sua formulação o argumento de Searle manteve sua força crítica? O argumento do quarto chinês pôs em marcha uma imensa indústria filosófica que gerou, nos anos seguintes, inúmeros contra-argumentos e objeções. Seria impossível revê-las aqui. Contudo, antes de avançar para o cerne deste artigo, examinarei a

---

<sup>1</sup> Bacharel em Filosofia pela USP, Mestre em Filosofia da Ciência pela UNICAMP e PhD pela University of Essex (Inglaterra), é um dos pioneiros da Filosofia da Mente no Brasil. Fez Pós-Doutorado nos Estados Unidos, sob orientação de Daniel Dennett, foi colaborador do Instituto de Estudos Avançados da USP e lecionou em várias Universidades brasileiras – como a UNESP, a UFSCar e a PUC-SP. Publicou 19 livros na área de Filosofia da Mente e Ciências Cognitivas. Mantém a página Filosofia da Mente no Brasil no Facebook (<https://www.facebook.com/filosofiadamentenobrasil/>). E-mail: [jteixe@terra.com.br](mailto:jteixe@terra.com.br) Lattes: <http://lattes.cnpq.br/8864985279295912>

objeção proposta por David Chalmers que, na minha percepção, é uma das mais convincentes. Exploro, em seguida, uma outra linha de objeções baseada na história da inteligência artificial que, recentemente desenvolveu máquinas híbridas com um hardware parcialmente biológico.

As primeiras máquinas semi-biológicas estão sendo desenvolvidas pelo pesquisador Kevin Warwick na Universidade de Reading, na Inglaterra. Essas máquinas híbridas desafiam o argumento do quarto chinês pelo fato de serem exemplares raros de uma interface entre o vivo e o mecânico-simbólico. Elas são, simultaneamente, uma forma de inteligência artificial simbólica e um ser vivo que preenche as condições necessárias para pensarem e serem conscientes. Com isso, a inteligência artificial simbólica se torna compatível com o naturalismo biológico proposto por Searle.

Warwick inverte a própria noção de ciborgue e propõe que, em vez de usar cérebros humanos para comandar máquinas pela adição de próteses, serão os robôs que se utilizarão de cérebros humanos para se tornar mais inteligentes. Em outras palavras, em vez de construir um cérebro para um robô, construiremos um robô para um cérebro humano.

Quando formulou seu argumento do quarto chinês, na década de 1980, Searle não podia imaginar a possibilidade da existência desses seres híbridos. E muito menos que um deles pudesse ser o ocupante do quarto chinês.

\*\*\*

O que é o argumento do quarto chinês?

Searle imaginou uma pessoa trancada em um quarto sem portas nem janelas, com apenas duas portinholas em paredes opostas. Essa pessoa só conhecia o português, mas lhe era fornecido um texto em chinês e uma espécie de tabela com regras e truques (escritos em português) para que ela, a partir de sentenças escritas em chinês, gerasse novas sentenças em língua chinesa. De vez em quando, abria-se uma das portinholas do quarto e um novo texto escrito em chinês lhe era fornecido.

O ocupante do quarto gerava um terceiro texto em chinês, com base no texto inicial escrito em chinês e nos novos textos que foram introduzidos, todos em chinês, usando as regras de transformação que estavam na tabela, escrita em português. Como o processo foi repetido regularmente, o ocupante do quarto adquiriu uma habilidade muito grande no manejo das regras de transformação.

Esse é, em linhas gerais, o argumento do quarto chinês. Searle nos diz que essa situação corresponde ao que ocorre no interior de um computador dotado de um programa para compreender histórias: o texto inicial, que está com a pessoa trancada no quarto, corresponde à história que é fornecida ao computador. As novas sentenças, geradas pela pessoa com base nas regras de transformação, podem muito bem ser as respostas do programa às perguntas que foram feitas com base no texto.

Ocorre que a pessoa que está no interior do quarto, manipulando a tabela com as regras de transformação, embora produza sentenças que são respostas adequadas às perguntas sobre o texto em chinês, *não compreende chinês*. Da mesma forma, um programa de computador apenas manipula símbolos – símbolos que não têm nenhum *significado* para a máquina. Trata-se de uma manipulação de símbolos inteiramente cega.

Por meio do argumento do quarto chinês Searle questionava se o famoso código binário de zeros e uns (0s e 1s) poderia dar origem ao significado e se ele não depende, inevitavelmente, de um intérprete humano. Como conceber uma máquina inteligente se ela não puder saber o significado do que ela faz? A barreira pode ser imensa, pois, para isso, ela precisaria ser uma máquina consciente, uma possibilidade descartada por Searle.

Searle afirmava que isso nunca seria possível, pois a consciência e, com ela, o significado, só podem ser produzidos por cérebros vivos e nunca por uma máquina composta de partes inanimadas. Se consciência e inteligência são exclusividade dos seres vivos ou, mais especificamente, de cérebros complexos como o do ser humano, seria preciso antes replicá-lo, para só depois produzir artificialmente a inteligência.

A seleção de informação relevante, a partir da qual é possível planejar para resolver problemas, pressupõe que ela seja significativa. O uso da informação, que a distingue da capacidade de simplesmente estocá-la, exige que se atribua a ela, previamente, um significado. Sem a atribuição de significado não é possível o *uso* da informação, componente essencial da inteligência. A questão é saber se essa atribuição de significado só é possível para seres vivos e, mais especificamente, para seres conscientes, como propõe Searle.

Do ponto de vista de um observador externo o quarto chinês é uma caixa preta, argumenta Searle. Mas, o que dizer de nosso cérebro? Ele também é uma caixa preta, pois não sabemos como ele processa a compreensão de textos. Tampouco sabemos sobre os processos que nos tornam seres conscientes. Estamos quase em pé de igualdade com o quarto no experimento mental de Searle, exceto pelo fato de sermos criaturas conscientes.

\*\*\*

Para Searle, a compreensão não é um processo intrínseco aos computadores digitais. Somos nós que atribuímos compreensão ao processo. Da mesma forma, no quarto chinês não há significado nos símbolos. O sistema pode se comportar sem nenhuma compreensão do que faz. Somos nós que atribuímos a ele a manipulação dos símbolos corretos em chinês. Podemos atribuir, metaforicamente, compreensão ao processo da mesma forma que atribuímos compreensão a um computador de mesa, mas não há nada de intrínseco ali. Para Searle, no caso dos computadores ou do quarto chinês não há nada de intrínseco porque é o observador que atribui ao sistema a compreensão do chinês. A chave para a compreensão está no mecanismo causal do cérebro e não no comportamento, de acordo com Searle.

A atenção de Searle para as propriedades causais do cérebro em detrimento da sintaxe formal de um programa é fruto de sua concepção da mente como um órgão biológico, um sistema com suas propriedades características. A mente é uma característica dos corpos humanos que evoluiu após milhões de anos, sendo extremamente útil pelas suas propriedades.

Toda a argumentação de Searle revela sua atenção para o que ele considera primordial no estudo da mente: a consciência. Todo o argumento do quarto chinês é proposto de maneira que a manipulação dos símbolos é feita sem nenhuma consciência por parte do ocupante do quarto. Para Searle, a consciência é parte fundamental do significado. Searle não admite que haja mente sem consciência. As propriedades semânticas estão estritamente ligadas à consciência.

Chalmers critica Searle por não distinguir entre sintaxe e implementação de sintaxe em seu argumento do quarto chinês. A sintaxe é a abstração formal de um processo, o que não acontece com o próprio processo implementado, que possui seus próprios poderes causais. A alegação de que um computador digital procede manipulando símbolos sem nenhum sentido é consequência da proposição de Searle de que a sintaxe é puramente formal e que é meramente simbólica mesmo quando implementada. Porém todo símbolo possui seus correspondentes em um substrato físico que procede causalmente dentro do sistema.

Dessa noção de sintaxe deriva grande parte da crítica de Searle ao modelo computacional da mente, que a concebe como um programa rodando em uma máquina. No entanto, como observa Chalmers, a sintaxe, quando implementada em um sistema, não é uma mera abstração, mas um processo que relaciona sua própria estrutura formal com as propriedades causais do sistema em que é implementada.

Simplificando o argumento de Chalmers contra Searle: uma receita de bolo não é a mesma coisa que um bolo. Searle critica as instruções contidas na receita como se ela fosse o

próprio bolo, esquecendo-se de que elas não são um bolo. Para que elas se tornem um bolo precisam ser implementadas, é preciso realizá-las em algum tipo de base material, no caso, os ingredientes. Em outras palavras, o argumento do quarto chinês é uma confusão categorial.

\*\*\*

A proposta da inteligência artificial no século XXI não é mais replicar a vida, mas associar-se a ela. Essa possibilidade é aberta pela *inteligência artificial biológica*. Ela mistura humanos com robôs a partir da cultura de neurônios *in vitro*. Sua proposta não é replicar a natureza, mas se apropriar dela e dar continuidade à evolução através de uma combinação entre seres vivos e próteses.

Na última década, as pesquisas de Kevin Warwick, professor da Universidade de Reading, na Inglaterra, impulsionaram a inteligência artificial biológica. Os primeiros passos foram dados usando neurônios de fetos de ratos em culturas que não ultrapassam as 100 mil unidades.

A cultura de um cérebro inicia-se com a coleta do tecido cortical por meio de enzimas. Esses neurônios passam a crescer em uma incubadora da qual eles recebem nutrientes e são mantidos em uma temperatura constante de 37 graus. Na medida em que os neurônios são alimentados, novas conexões entre eles vão surgindo, e há um aumento de sua comunicação química e elétrica.

Esses neurônios cultivados *in vitro* são, em seguida, conectados a um robô *miabot* que lhes fornece informação sobre seu entorno. Isso é feito ligando os neurônios a um conjunto de microelétrodos através de uma conexão sem fio Bluetooth, estabelecendo uma interface elétrica com o robô, que permite controlar seus movimentos. Comportamentos bem-sucedidos, que tendem a se repetir, aumentam a intensidade e a frequência dos impulsos elétricos recebidos pelo robô e são transmitidos para a cultura, que passa a ser um cérebro na proveta. Esses impulsos criam memórias, cuja transcrição cerebral é o aumento de conectividade entre algumas populações de neurônios.

É possível monitorar os outputs elétricos desse cérebro na proveta e enviar-lhe informação sensorial. Os outputs podem, então, ser utilizados para mover adequadamente o robô. A relação input/output estabelece uma realimentação mútua entre o cérebro cultivado e os mecanismos do robô que, com o tempo, faz com que, além de reforçar conexões, novas camadas de neurônios se formem sobre os microelétrodos.

A perspectiva futura é a utilização de neurônios humanos nessas culturas. Eles poderão ser obtidos por meio das pessoas que façam a doação de seus cérebros após a morte. Se for doada uma quantidade pequena de neurônios, será necessário esperar vários anos até que as culturas se desenvolvam e se transformem em um cérebro humano adulto. Há também a possibilidade de doação de um cérebro humano intacto e, nesse caso, é muito provável que muitas informações e experiências estocadas na memória estejam preservadas em suas conexões neuronais. Se seus neurônios e suas conexões forem adequadamente preservados e cultivados em um cérebro na proveta, todos esses dados serão herdados.

A conexão desse cérebro com um corpo robótico poderá ser feita se, antes da morte, o doador do cérebro autorizar a introdução de milhões de nano-eléttodos em seu corpo, através da artéria femoral, como ocorre na realização de uma angiografia, um exame no qual um contraste é injetado para tornar visíveis os ramos capilares de um órgão. Os nano-eléttodos subirão para o cérebro e se instalarão nos capilares. Com isso, será possível detectar a atividade elétrica de cada neurônio, gerando a informação necessária para mover adequadamente o corpo robótico acoplado a esse cérebro após a sua morte e remoção para uma proveta.

O projeto de inserir nano-eléttodos no cérebro está sendo desenvolvido pelo neurocientista colombiano Rodolfo Llinás. Não há risco de que esses nano-elettodos penetrem no cérebro, o que poderia danificá-lo. Os capilares do cérebro estão muito próximos uns dos outros, formando uma barreira sanguínea que impede a penetração de quaisquer moléculas maiores do que as de oxigênio, glicose e álcool. É por isso que, se injetarmos um líquido de contraste na veia de uma pessoa, ele se espalhará por todo o corpo, mas não conseguirá penetrar no cérebro. Inversamente, se um contraste for aplicado no cérebro, ele não se espalhará pelo resto do corpo.

Embora a atividade elétrica dos neurônios seja contínua, os nano-eléttodos a transformarão em uma sucessão de estados discretos, gerando pulsos, ou seja, 0s e 1s na forma de um código que poderá ser decifrado por um supercomputador, permitindo a comunicação com o cérebro na proveta. Além disso, a interação do cérebro na proveta com o meio ambiente por meio da movimentação do corpo robótico permitirá novos aprendizados e uma constante atualização de dados acerca de seu entorno. É possível também que, através de interfaces cérebro-máquina, como as que estão sendo desenvolvidas pelo neurocientista brasileiro Miguel Nicolelis (2011), o cérebro na proveta possa se comunicar diretamente usando o pensamento para acionar partes do corpo.

Será que poderíamos atribuir a esse robô consciência, compreensão e percepção do significado? A principal conclusão do argumento do quarto chinês de Searle é a de que a consciência exige um cérebro biológico, o que, até agora, não foi possível simular com chips de silício. Mas o que dizer de um robô com um cérebro que poderá ser idêntico ao de um ser humano e ao qual foi acoplado um corpo artificial?

O que ocorre nesse caso é o acoplamento de um cérebro na proveta a uma máquina, que passa a ser seu corpo. Em vez de trocarmos os neurônios por chips, como já foi proposto por alguns pesquisadores, são as partes do corpo que vão sendo substituídas por equivalentes robóticos até o ponto de só restarem próteses, com exceção do cérebro.

Será que uma máquina desse tipo não preservaria os poderes causais do cérebro a ponto de gerar estados intencionais e produzir significado? Em outras palavras, se essa máquina for construída, não teríamos uma resposta, não apenas conceitual, mas empírica, para o argumento do quarto chinês?

Dessa perspectiva, temos uma resposta para o argumento do quarto chinês. Searle defende que a substituição dos neurônios por chips não permite a manutenção da consciência. Talvez essa conclusão seja verdadeira. No entanto, ela não é suficiente para descartar a possibilidade da reprodução da consciência pelo cultivo de neurônios humanos. Se o ocupante do quarto for um cérebro na proveta, o argumento do quarto chinês cai por terra sem que seja necessário contradizer a ideia de Searle de que a consciência depende da matéria viva.

É possível argumentar, contra essa ideia, que a mente não está somente no cérebro, mas espalhada pelo corpo, o que não acontece no caso de um robô com partes mecânicas acopladas a um cérebro na proveta. No entanto, esse argumento pode levar a consequências paradoxais. Imagine uma pessoa que tenha perdido sua mão, mas que consegue levá-la para um hospital no qual há um cirurgião especializado em reimplantes. Se a mente está espalhada pelo corpo, o cirurgião terá de anestésiar, não apenas a pessoa, como também sua mão, antes de reimplantá-la. Mas, como ironizou Dennett no artigo “Using reductio ad absurdum” (Usando a redução ao absurdo, 2013), essa é uma conclusão bizarra. Ninguém sente dor em uma mão que se separou de seu corpo.

A possibilidade de cultivar um cérebro na proveta ainda é hipotética. No entanto, ela abre a perspectiva de desenvolver culturas de neurônios humanos cujo número se aproxime ao do nosso cérebro ou até o supere. Nesse caso, será preciso modificar o código genético dos neurônios para que eles continuem se desenvolvendo após atingir a idade madura, sem que o cérebro ampliado perca suas características anatômicas. Robôs com forma humanoide contendo



cérebros desenvolvidos *in vitro* serão as novas superinteligências produzidas pela inteligência artificial biológica. Com eles, serão criadas unidades organismo-máquina que poderão superar a capacidade de processamento de informação do cérebro humano e serem mais inteligentes do que nós.

O cérebro na proveta desafia o convívio entre a filosofia e a tecnologia ao sugerir que a subjetividade e a consciência não se contrapõem à visão científica e tecnológica do mundo apesar de não ter sido possível, até hoje, resolver o problema mente-cérebro de um ponto de vista metafísico. Uma máquina pode incorporar subjetividade e consciência tomando-as emprestadas de um ser humano e, com isso, gerar e reconhecer o significado.

A inteligência artificial biológica abrirá novas perspectivas tecnológicas, filosóficas e antropológicas, unificando em um só paradigma a ciência da computação e a neurociência. Essa abordagem unificada mostra, também, que a inteligência artificial é, pelo menos *conceitualmente* possível, o que permite descartar algumas das objeções clássicas contra a sua realização, como é o caso do argumento do quarto chinês, de Searle.

\*\*\*

## Referências

CHALMERS, D. *The Conscious Mind*. Oxford: Oxford University Press, 1996.

DENNETT, D. Reductio ad Absurdum. In: *Intuition Pumps*. New York: W.W. Norton & Company, 2013.

KURZWEIL, R. *A era das máquinas espirituais*. São Paulo: Aleph, 2007. (Originalmente publicado em 1999).

LLINÁS, R. *Wires in the Brain*. Disponível em: <https://www.youtube.com/watch?v=JzrrLNI2n44>. Acesso em: julho 2020.

NICOLELIS, M. *Muito além do nosso eu*. São Paulo: Companhia das Letras, 2011.

PRESTON, J. & BISHOP, M. *Views into the Chinese Room*. Oxford: Oxford University Press, 2002.

SEARLE, J. I Married a Computer. In: RICHARDS, J. W (Ed.). *Are We Spiritual Machines?* Seattle: Discovery Institute, 2002.



SEARLE, J. *A redescoberta da mente*. São Paulo: WMF Martins Fontes, 1997 (Coleção Tópicos).

SEARLE, J. Mentes, cérebros e programas. Disponível em <http://opessoa.fflch.usp.br/sites/opessoa.fflch.usp.br/files/Searle-Port-2.pdf> Acesso em: setembro/2020.

TEIXEIRA, J. de F. *Mente, cérebro e cognição*. Petrópolis: Vozes, 2000.

TEIXEIRA, J. de F. *O cérebro e o robô*. São Paulo: Editora Paulus, 2015.

WARWICK, K. *Artificial Intelligence: The Basics*. Londres: Routledge, 2012.

WARWICK, K. Robots with Biological Brains. In: LIN, P.; ABNEY, K. & BEKEY, G. (Eds.). *Robot Ethics*. Cambridge: The MIT Press, 2012.

WARWICK, K. *March of the Machines*. London: Century, 1997.