Revista Eletrônica de Filosofia
Philosophy Eletronic Journal
ISSN 1809-8428

São Paulo: Centro de Estudos de Pragmatismo
Programa de Estudos Pós-Graduados em Filosofia
Pontifícia Universidade Católica de São Paulo
Disponível em http://www.pucsp.br/pragmatismo

Vol. 12, nº. 1, janeiro-junho, 2015, p. 157-169

UMA MÁQUINA PODERIA PENSAR?
É improvável que a IA clássica produza máquinas conscientes, mas sistemas que imitam o cérebro talvez consigam*

Paul M. Churchland e Patricia Smith Churchland**

Tradução de Nara Ebres Bachinski
Universidade Federal de Santa Maria – RS
naraebresb@gmail.com

A pesquisa em inteligência artificial está passando por uma revolução. Para explicar como e por que, e para colocar o argumento de John R. Searle em perspectiva, primeiro precisamos de um *flashback*.

No início da década de 1950, a questão antiga e vaga "Uma máquina poderia pensar?" foi substituída por questões mais acessíveis: "Uma máquina que manipulasse símbolos físicos de acordo com regras sensíveis a estruturas [*structure-sensitive*] poderia pensar?". Esta questão foi um avanço, porque a lógica formal e a teoria computacional tinham visto grandes desenvolvimentos no meio século precedente. Os teóricos passaram a apreciar o enorme poder dos sistemas simbólicos abstratos que passam por transformações governadas por regras. Se esses sistemas pudessem ser automatizados, então seu poder computacional abstrato, ao que parece, seria exibido em um sistema físico real. Esse *insight* gerou um programa de pesquisa bem definido e com bases teóricas profundas.

---

* Originalmente publicado em *Scientific American* 262.1 (1990): 32-37. Tradução publicada com permissão dos editores. Tradução revisada por Rogério Passos Severo.

** Paul M. Churchland e Patricia Smith Churchland são professores de Filosofia na Universidade da Califórnia em San Diego. Juntos têm estudado a natureza da mente e do conhecimento há duas décadas. Paul Churchland concentra-se na natureza do conhecimento *científico* e seu desenvolvimento, enquanto Patricia Churchland concentra-se nas neurociências e em como o cérebro sustenta a cognição. *Matéria e Consciência*, de Paul Churchland é um livro-texto *standard* em Filosofia da Mente, e *Neurofilosofia*, de Patricia Churchland, reúne teorias da cognição da Filosofia e da Biologia. Paul Churchland é atualmente chefe do departamento de Filosofia da UCSD, e os dois são, respectivamente, presidente e ex-presidente da Society for Philosophy and Psychology. Patricia Churchland também é professora no Salk Institute for Biological Studies em San Diego. Os Churchlands são também membros do corpo docente de ciências cognitivas da UCSD e de seu Institute for Neural Computation e de seu programa de Science Studies.

Uma máquina poderia pensar? Havia muitas razões para dizer que sim. Uma das razões primeiras e mais profundas estava em dois resultados importantes na teoria da computação. O primeiro foi a tese de Church, que afirma que toda função efetivamente computável é recursivamente computável. Efetivamente computável significa que existe um processo "automático" para determinar, em tempo finito, o *output* da função para um dado *input*. Recursivamente computável significa, mais especificamente, que existe um conjunto finito de operações que podem ser aplicadas a um dado *input*, e, em seguida, aplicada de novo e de novo aos sucessivos resultados de tais aplicações, para gerar o *output* da função em tempo finito. A noção de um procedimento automático é intuitiva e não formal; assim, a tese de Church não admite uma prova formal. Mas ela vai ao cerne do que é para computar, e muitas linhas de evidência convergem para apoiá-la.

O segundo resultado importante foi a demonstração por Alan M. Turing que qualquer função recursivamente computável pode ser computada, em um tempo finito, por um tipo de máquina manipuladora de símbolos maximamente simples, que veio a ser chamada de "máquina de Turing universal". Essa máquina é guiada por um conjunto de regras recursivamente aplicáveis que são sensíveis à identidade, à ordem e à disposição dos símbolos elementares que encontra como *input*.

Estes dois resultados implicam algo notável, ou seja, que um computador digital padrão – dados apenas o programa adequado, uma memória suficientemente grande e tempo suficiente – pode computar *qualquer* função *input-output* governada por regras. Isto é, pode exibir qualquer padrão sistemático de respostas ao ambiente.

Mais especificamente, estes resultados implicam que uma máquina manipuladora de símbolos adequadamente programada (a seguir, máquina MS) deve ser capaz de passar no teste de Turing para inteligência consciente. O teste de Turing é um teste puramente comportamental para a inteligência consciente, mas, mesmo assim, é um teste muito exigente. (Se trata-se de um teste justo, isso será abordado abaixo, onde também encontraremos outro "teste", bem diferente, para a inteligência consciente.) Na versão original do teste de Turing, os *inputs* para a máquina MS são questões e observações conversacionais digitadas em um console por mim ou você, e os *outputs* são respostas digitadas pela máquina MS. A máquina passa nesse teste de consciência inteligente se suas respostas não puderem ser discriminadas das respostas digitadas por uma pessoa real e inteligente. É claro, atualmente niguém conhece a função que produziria o *output* comportamental de uma pessoa consciente. Mas os resultados de Church e Turing asseguram-nos que, qualquer que seja essa função (presumivelmente efetiva), uma máquina MS adequada poderia computá-la.

Essa é uma conclusão significativa, especialmente porque o retrato que Turing fez de uma interação puramente teledigitada é uma restrição desnecessária. A mesma conclusão segue-se mesmo se a máquina MS interaja com o mundo de formas mais complexas: por visão direta, discurso real e assim por diante. Afinal, uma função recursiva mais complexa ainda é Turing-computável. O único problema restante é o de identificar a função indubitavelmente complexa que governa o padrão humano de resposta ao ambiente e, em seguida, escrever o programa (o conjunto de regras recursivamente aplicáveis) pelo qual a máquina MS irá computá-la. Esses objetivos constituem o programa de pesquisa fundamental da IA clássica.

Os resultados iniciais foram positivos. Máquinas MS com programas inteligentes realizaram uma variedade de atividades ostensivamente cognitivas. Elas responderam a instruções complexas, resolveram problemas aritméticos, algébricos e táticos complexos, jogaram damas e xadrez, provaram teoremas e engajaram-se em diálogos simples. Os desempenhos continuaram melhorando com o aparecimento de memórias maiores e máquinas mais rápidas, e com a utilização de programas mais longos e mais espertos. A IA clássica, ou "escrevedora de programas", foi um esforço de investigação vigoroso e bem sucedido em quase todas as perspectivas. A negação ocasional que uma máquina MS poderá, um dia, pensar pareceu desinformada e mal motivada. O argumento em favor de uma resposta positiva à pergunta do título deste artigo foi esmagadora.

Havia alguns problemas, é claro. Em primeiro lugar, admitidamente, máquinas MS não eram muito parecidas com cérebros. Mesmo aqui, no entanto, a abordagem clássica teve uma resposta convincente. Primeiro, o material físico de qualquer máquina MS não tem essencialmente nada a ver com a função que ela computa. Isso é fixado pelo seu programa. Em segundo lugar, os detalhes de engenharia da arquitetura funcional de qualquer máquina também são irrelevantes, uma vez que diferentes arquiteturas que rodam programas bem diferentes podem ainda assim estar computando a mesma função *input-output*.

De acordo com isso, a IA procurou encontrar a função *input-output* característica da inteligência e o mais eficiente dos muitos programas possíveis para computá-la. A maneira idiossincrática pela qual o cérebro computa a função simplesmente não importa, dizia-se. Isso completa a justificativa para a IA clássica e para uma resposta positiva à nossa pergunta do título.

Uma máquina poderia pensar? Havia também alguns argumentos para dizer não. Durante a década de 1960, argumentos negativos interessantes eram relativamente raros. Uma objeção feita ocasionalmente foi de que o pensamento era um processo não físico em uma alma imaterial. Mas essa resistência dualista não era nem evolutivamente nem explicativamente plausível. Ela teve um impacto negligenciável na pesquisa de IA.

Uma linha bem diferente de objeção foi mais bem sucedida em ganhar a atenção da comunidade de IA. Em 1972, Hubert L. Dreyfus publicou um livro que foi altamente crítico das simulações ostensivas da atividade cognitiva. Ele argumentou que eram inadequadas como simulações de cognição genuína e indicou um padrão de falha nessas tentativas. O que estaria faltando, ele sugeriu, era a grande quantidade de conhecimentos inarticulados pressupostos que cada pessoa possui, e a capacidade de senso comum de ater-se aos aspectos relevantes daqueles conhecimentos à medida que as circunstâncias em mutação exigem. Dreyfus não negou a possibilidade de que um sistema físico artificial de algum tipo poderia pensar, mas ele foi muito crítico em relação à ideia de que isso pode ser alcançado apenas por manipulação de símbolos nas mãos de regras recursivamente aplicáveis.

As queixas de Dreyfus foram amplamente percebidas, na comunidade da IA e no âmbito da disciplina de Filosofia, como míopes e antipáticas, como remetendo às simplificações inevitáveis de um esforço de pesquisa ainda em sua juventude. Estes déficits podem ser reais, mas com certeza eles eram temporários. Máquinas maiores e programas melhores deveriam repará-los no momento oportuno. O tempo, sentia-

se, estava do lado da IA. Aqui, novamente, o impacto sobre a pesquisa foi negligenciável.

O tempo também estava do lado de Dreyfus: a taxa de retorno cognitivo, relativamente ao aumento da velocidade e da memória, começou a diminuir no final dos anos 1970 e início dos anos 1980. A simulação de reconhecimento de objetos no sistema visual, por exemplo, mostrou-se computacionalmente intensa em um grau inesperado. Resultados realistas exigiam cada vez mais tempo dos computadores, períodos de tempo que em muito excediam o que é exigido por um sistema visual real; essa lentidão relativa das simulações era obscuramente curiosa; a propagação de sinais em um computador é aproximadamente um milhão de vezes mais rápido que no cérebro, e a frequência do processador central de um computador é maior do que qualquer frequência encontrada no cérebro por uma margem similarmente dramática.

Além disso, um desempenho realista exigia que o programa de computador tivesse acesso a uma base de conhecimento extremamente grande. Construir a base de conhecimento relevante era problema suficiente, e foi agravado pelo problema de como acessar apenas as partes contextualmente relevantes daquela base de conhecimento, em tempo real. À medida que a base de conhecimento ficava maior e melhor, o problema do acesso piorou. A busca exaustiva levava tempo demais, e a heurísticas da relevância evoluiu pouco. Preocupações do tipo que Dreyfus tinha apontado finalmente começaram a se firmar aqui e ali, mesmo entre pesquisadores de IA.

Nessa época (1980), John Searle escreveu uma crítica nova e bem diferente à suposição mais básica do programa de pesquisa clássico: a ideia de que a manipulação apropriada de símbolos estruturados, pela aplicação recursiva de regras estrutura-sensível, poderia constituir inteligência consciente.

O argumento de Searle baseia-se em um experimento de pensamento que exibe duas características cruciais. Em primeiro lugar, ele descreve uma máquina de SM que realiza, supõe-se, uma função de *input-output* adequada para sustentar uma conversa bem sucedida no teste de Turing, conduzido inteiramente em chinês. Em segundo lugar, a estrutura interna da máquina é tal que, não importa como ela se comporte, um observador permanece certo de que nem a máquina nem qualquer parte dela compreende chinês. Tudo que ela contém é um falante monolíngue do Inglês que segue um conjunto escrito de instruções para manipular os símbolos chineses, que chegam e saem através de uma pequena abertura. Em suma, o sistema deve passar no teste Turing, embora o próprio sistema não tenha qualquer compreensão genuína de chinês ou conteúdo semântico chinês real [cf. “É a mente do cérebro um programa de computador?”, de John R. Searle (disponível em português em: L. BonJour; A. Baker (ed.) *Filosofia: textos fundamentais comentados*. Porto Alegre: Artmed, 2010)].

A lição geral extraída é que qualquer sistema que meramente manipule símbolos físicos de acordo com regras sensíveis às estruturas será, na melhor das hipóteses, uma imitação vazia da inteligência consciente real, porque é impossível gerar uma “semântica real” apenas operando uma “sintaxe vazia”. Aqui, devemos salientar, Searle está impondo um teste não comportamental para a consciência: os elementos de inteligência consciente devem possuir conteúdo semântico real.

Somos tentados a reclamar que a experiência de pensamento de Searle é injusta, porque o seu sistema de Rube Goldberg computará com uma lentidão

absurda. Searle insiste, porém, que a velocidade é rigorosamente irrelevante aqui. Um pensador lento ainda deve ser um pensador real. Tudo que é essencial para a duplicação do pensamento, segundo a IA clássica, é dito estar presente no quarto chinês.

O artigo de Searle provocou uma reação animada de pesquisadores de IA, bem como de psicólogos e filósofos. Em geral, no entanto, ele encontrou uma recepção ainda mais hostil do que Dreyfus experimentara. Em seu artigo, que acompanha o nosso nesta edição, Searle enumera de modo franco várias dessas respostas críticas. Pensamos que muitas delas são razoáveis, especialmente aquelas que "não fogem da raia", insistindo que, embora seja terrivelmente lento, o sistema geral sala-mais-conteúdos, de fato, entende chinês.

Pensamos que essas são boas respostas, mas não porque pensamos que oquarto compreende chinês. Concordamos com Searle que esse não é o caso. Em vez disso, são respostas boas, porque refletem uma recusa em aceitar o terceiro axioma crucial do argumento de Searle: "*A sintaxe por si só não é constitutiva e nem suficiente para a semântica*". Talvez esse axioma seja verdadeiro, mas Searle não pode legitimamente fingir saber que é. Além disso, supor a sua verdade é cometer petição de princípio contra o programa de pesquisa da IA clássica, pois esse programa baseia-se na suposição muito interessante de que, se pudéssemos colocar em movimento uma dança interna adequadamente estruturada de elementos sintáticos, devidamente ligada a *inputs* e *outputs*, isso pode produzir os mesmos estados cognitivos e realizações encontradas em seres humanos.

A petição de princípio do axioma 3 de Searle fica clara quando comparada diretamente a sua conclusão 1: "*Os programas não são nem constitutivos de nem suficientes para mentes*" Claramente, seu terceiro *axioma* já está carregando 90%do peso dessa conclusão quase idêntica. É por isso que o experimento de pensamento de Searle dedica-se a fortalecer especificamente o axioma 3. Esse é o ponto do quarto chinês.

Embora a história do quarto chinês torne o axioma 3 tentador para os incautos, não pensamos que ele é bem sucedido em estabelecer o axioma 3, e oferecemos, abaixo, um argumento paralelo para ilustraçar o seu fracasso. Uma única instância, manifestamente falaciosa, de um argumento contestado muitas vezes proporciona muito mais *insight* do que um livro cheio de distinções lógica.

O estilo de ceticismo de Searle tem amplo precedente na história da ciência. O bispo irlandês do século XVIII, George Berkeley, achava ininteligível que a compressão de ondas no ar, por si mesmas, pudessem constituir ou ser suficientes para o som objetivo. O poeta-artista inglês, William Blake, e o poeta-naturalista alemão, Johann W. von Goethe, achavam inconcebível que pequenas partículas, por si mesmas, pudessem constituir ou serem suficientes para o fenômeno objetivo da luz. Mesmo neste século, houve pessoas que achavam inimaginável que a matéria inanimada, por si mesma, e como quer que tivesse organizada, pudesse constituir ou ser suficiente para a vida. Claramente, o que as pessoas podem ou não podem imaginar muitas vezes não tem nada a ver com o que é ou não é o caso, mesmo que as pessoas envolvidas sejam muito inteligentes.

Para ver como essa lição aplica-se ao caso de Searle, considere um argumento deliberadamente fabricado, paralelo ao dele e ao experimento de pensamento que o apoia.

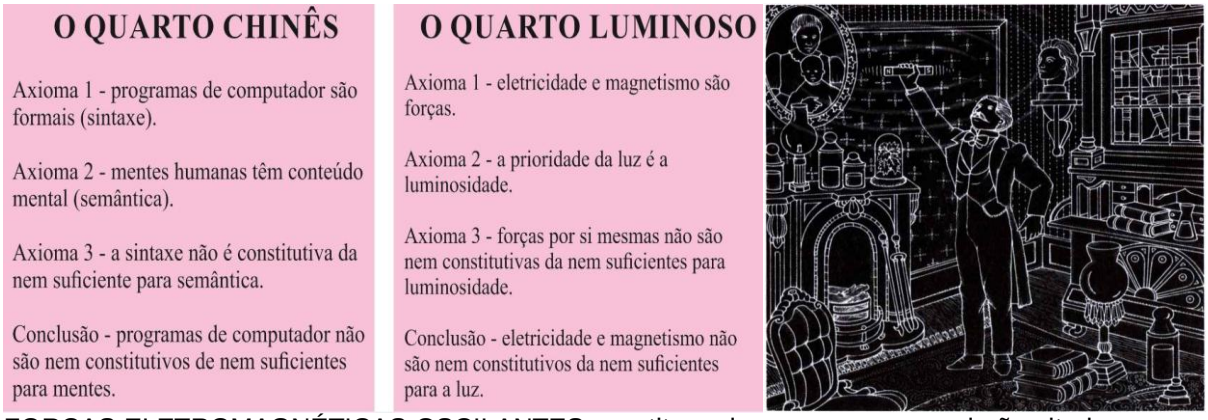
Axioma 1. *Eletricidade e magnetismo são forças.*

Axioma 2. *A propriedade essencial da luz é a luminosidade*.

Axioma 3. *Forças por si mesmas não são nem constitutivas da nem suficientes para a luminosidade.*

Conclusão. *Eletricidade e magnetismo não são nem constitutivos da nem suficientes para a luz.*

Imagine esse argumento sendo apresentado logo após a sugestão de 1864, feita por James Clerk Maxwell, em que a luz e as ondas eletromagnéticas são idênticas, mas antes da apreciação plena de todo o mundo plenamente reconhecer os paralelos sistemáticos entre as propriedades da luz e as propriedades das ondas eletromagnéticas. Esse argumento poderia ter servido como objeção convincente para a hipótese imaginativa de Maxwell, especialmente se fosse acompanhada pelo seguinte comentário em apoio axioma 3.



FORÇAS ELETROMAGNÉTICAS OSCILANTES constituem luz mesmo que um imã agitado por uma pessoa pareça não produzir luz alguma. Da mesma forma, manipulação de símbolos baseado em regras poderia constituir inteligência mesmo que o sistema baseado em regras dentro do "Quarto Chinês", de John R. Searle, pareça carecer de compreensão real.

> Considere um quarto escuro contendo um homem segurando um imã ou um objeto carregado eletricamente. Se o homem agita o ímã para cima e para baixo, então, de acordo com a teoria da luminosidade artificial (LA) de Maxwell, ele irá iniciar um círculo de propagação de ondas eletromagnéticas e assim ficará luminoso. Mas, como todos nós, que já brincamos com ímãs ou bolas carregadas eletricamente, sabemos bem, as suas forças (ou quaisquer outras forças), mesmo quando colocadas em movimento, não produzem nenhuma luminosidade. É inconcebível que se poderia constituir luminosidade real apenas movendo forças!

Como deveria Maxwell responder a este desafio? Ele poderia começar insistindo que o experimento do "quarto luminoso" é uma exposição enganosa do fenômeno da luminosidade, porque a frequência de oscilação do ímã é absurdamente baixa, baixa demais por um fator de 10^{15}. Isso poderia muito bem produzir a resposta impaciente de que a frequência não tem nada a ver com isso, que o quarto com o ímã balançando já contém tudo o que é essencial para a luz, de acordo com a própria teoria de Maxwell.

Em resposta, Maxwell poderia assumir o desafio e afirmar, com razão, que o quarto realmente está banhado em luminosidade, embora em um grau ou qualidade demasiadamente fracos para ser apreciada. (Dada a baixa frequência com a qual o homem pode oscilar o ímã, o comprimento de onda das ondas eletromagnéticas produzidas é muito longo e sua intensidade é muito fraca para que retinas humanas reajam a elas.) Mas, no clima de entendimento aqui contemplado – na década de 1860 – essa tática provavelmente suscitaria risos e vaias. “Quarto luminoso? Nada disso, Sr. Maxwell. Está totalmente escuro lá dentro!”.

Infelizmente, o pobre Maxwell não tem rota escapatória fácil para sair dessa armadilha. Tudo o que ele pode fazer é insistir nos três pontos seguintes. Primeiro, o axioma 3 do argumento acima é falso. Na verdade, ele comete petição de princípio, apesar de sua plausibilidade intuitiva. Em segundo lugar, o experimento do quarto luminoso não demonstra nada interessante sobre a natureza da luz. E, em terceiro lugar, o que é necessário para resolver o problema da luz e da possibilidade de luminosidade artificial é um programa de investigação em curso para determinar se, sob as condições adequadas, o comportamento de ondas eletromagnéticas, de fato, reflete perfeitamente o comportamento da luz.

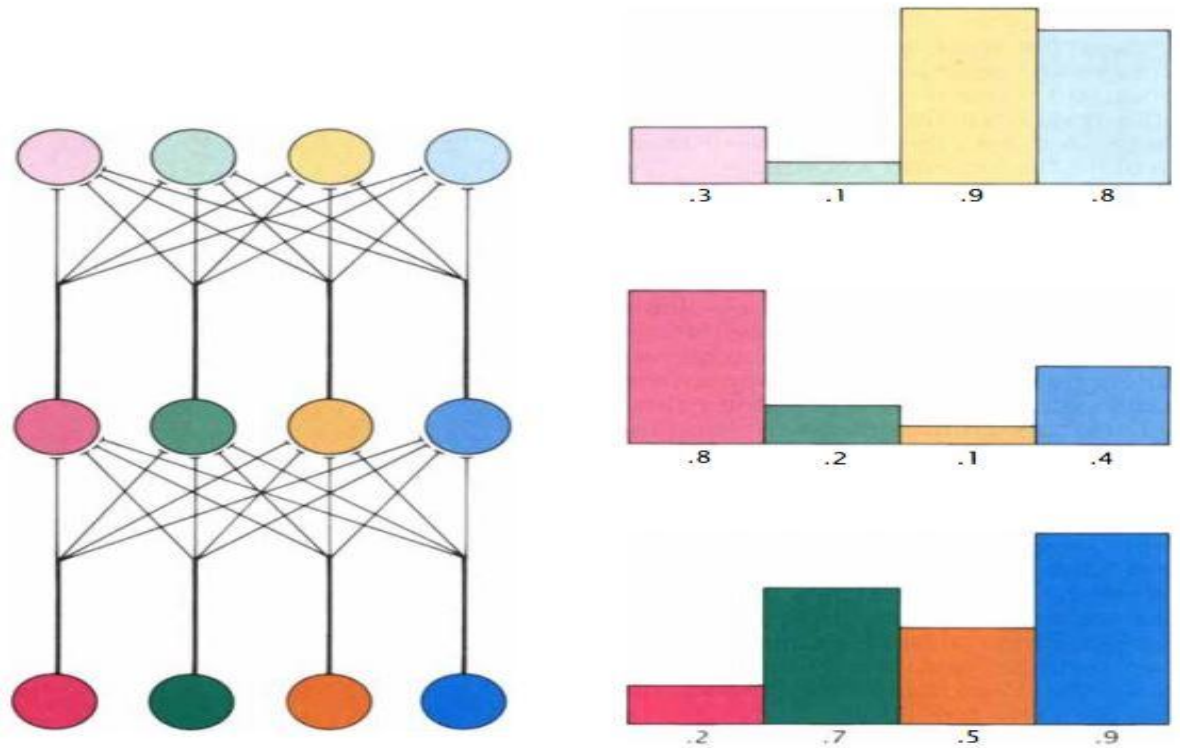
Esta também é a resposta que IA clássica deve dar ao argumento de Searle. Embora o quarto chinês de Searle possa parecer “semanticamente escuro”, ele não está em posição de insistir, baseado na força dessa aparência, que a manipulação de símbolos, governada por regras, jamais poderá constituir fenômeno semântico – especialmente quando as pessoas têm apenas uma compreensão de senso comum desinformado dos fenômenos semânticos e cognitivos que precisam ser explicados. Em vez de explorar nossa compreensão dessas coisas, o argumento de Searle explora livremente nossa ignorância deles.

Com essas críticas ao argumento de Searle, voltamos à questão de saber se o programa de pesquisa da IA clássica tem uma chance realista de resolver o problema da inteligência consciente e de produzir uma máquina que pense. Acreditamos que os prospectos são pobres, mas apoiamos essa opinião em razões muito diferentes das de Searle. Nossas razões derivam de falhas de desempenho específicas do programa de pesquisa da IA clássica, de uma variedade de lições aprendidas com o cérebro biológico e de uma nova classe de modelos computacionais inspirados na sua estrutura. Já indicamos algumas das falhas da IA clássica com respeito a tarefas que o cérebro executa com rapidez e eficiência. O consenso emergente sobre essas falhas é que a arquitetura funcional de máquinas MS clássicas são, simplesmente, a arquitetura errada para as tarefas muito exigentes que precisam ser executadas.

O que precisamos saber é isto: Como o cérebro consegue cognição? A engenharia reversa é uma prática comum na indústria. Quando uma nova tecnologia chega ao mercado, os concorrentes descobrem como ela funciona desmontando-a, e inferem a sua lógica estrutural. No caso do cérebro, esta estratégia apresenta um desafio incomumente duro, pois o cérebro é a coisa mais complicada e sofisticada no planeta. Mesmo assim, as neurociências têm revelado muito sobre o cérebro em uma grande variedade de níveis estruturais. Três pontos anatômicos proporcionarão um contraste básico com a arquitetura dos computadores eletrônicos convencionais.

Em primeiro lugar, os sistemas nervosos são máquinas em paralelo, no sentido de que os sinais são processados em milhões de diferentes vias, simultaneamente. A retina, por exemplo, apresenta o seu complexo de entrada para

o cérebro não em blocos de 8, 16, ou 32 elementos, como num computador de mesa, mas sim sob a forma de quase um milhão de elementos distintos de sinais, que chegam simultaneamente ao alvo do nervo ótico (o núcleo geniculado lateral), para então ser processado coletivamente, simultaneamente e de uma só vez. Em segundo lugar, a unidade de processamento básica do cérebro, o neurônio, é relativamente simples. Além disso, a sua resposta a sinais de entrada é analógica, não digital, na medida em que a sua freqüência de saída varia continuamente com os seus sinais de entrada. Em terceiro lugar, os axônios cerebrais, que se projetam de uma população neuronal a outra, são frequentemente acompanhados por axônios que retornam de sua população-alvo. Essas projeções descendentes ou recorrentes permitem ao cérebro modular o caráter de seu processamento sensorial. Mais importante ainda, sua existência torna o cérebro um verdadeiro sistema dinâmico, cujo comportamento contínuo é altamente complexo e, até certo ponto, independente de seus estímulos periféricos.



AS REDES NEURAIS modelam um aspecto central da microestrutura do cérebro. Nessa rede de três camadas, neurônios de *input* (canto inferior **esquerdo**) processam um padrão de ativações (canto inferior **direito**) e o repassam através de conexões ponderadas para uma camada oculta. Elementos na camada oculta somam seus muitos *inputs* e produzem um novo padrão de ativações. Este é repassado à camada de *output*, que realiza transformações adicionais. Ao todo, a rede transforma qualquer padrão de *input* em um padrão correspondente de *output*, tal como ditado pelo arranjo e força das muitas conexãos entre os neurônios.

Modelos de redes altamente simplificados têm sido úteis para sugerir como as redes neurais reais podem funcionar e também para revelar as propriedades computacionais de arquiteturas paralelas. Por exemplo, considere um modelo de três camadas constituído por unidades, parecidas com neurônios, totalmente conectadas, por conecções parecidas com axiônios, a unidades da próxima camada.

Um estímulo *input* produz algum nível de ativação em uma determinada unidade de *input*, que transmite um sinal de força proporcional ao longo de seu "axônio" para suas muitas conexões sinápticas, para as unidades ocultas. O efeito global é que um padrão de ativações, no outro lado do conjunto de unidades de *input*, produz um padrão distinto de ativações, no outro lado do conjunto de unidades ocultas.

A mesma história aplica-se às unidades de *output*. Como antes, um padrão de ativação nas unidades ocultas produz um padrão de ativação distinta nas unidades de *output*. Ao todo, essa rede é um dispositivo que permite transformar, qualquer um, de um grande número possível de vetores de *inputs* (padrões de ativação) em um vetor de *output* de correspondencia única. Trata-se de um dispositivo para computar uma função específica. Exatamente qual função ele computa é fixada pela configuração global dos seus pesos sinápticos.

Há vários procedimentos para ajustar os pesos, de modo a se obter uma rede que compute quase qualquer função, isto é, qualquer transformação vetor-para-vetor de que se possa desejar. Na verdade, pode-se até mesmo impor sobre ela uma função que não se é capaz de especificar, contanto que se possa fornecer um conjunto de exemplos de pares de *inputs* e de *outputs* desejados. Esse processo, chamado "treinando a rede", procede por ajustes sucessivos dos pesos na rede até que ele execute as transformações de *input-output* desejados.

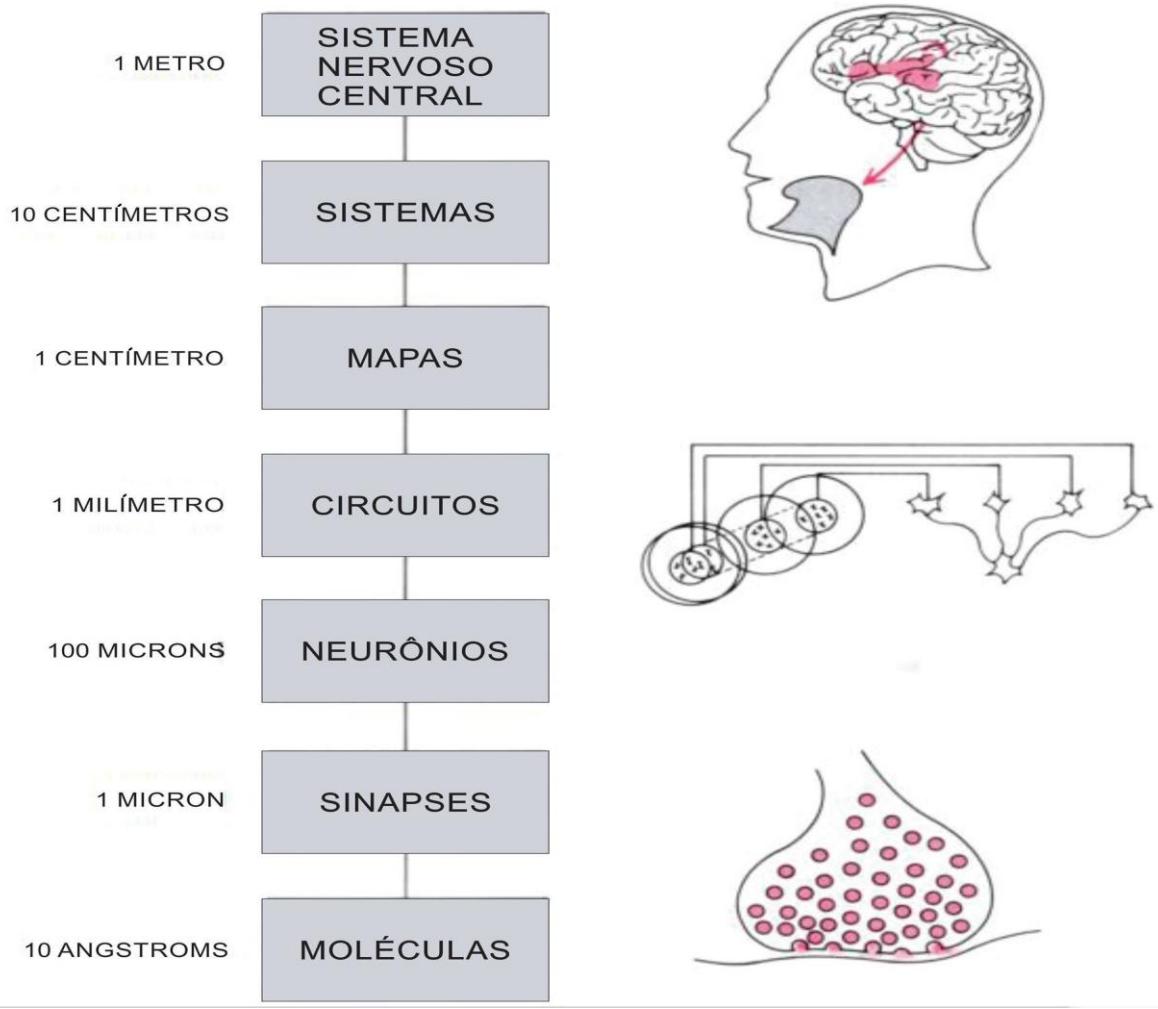
Embora, esse modelo de rede simplifique enormemente a estrutura do cérebro, ele ilustra várias ideias importantes. Em primeiro lugar, uma arquitetura paralela fornece uma vantagem dramática de velocidade sobre um computador convencional, pois as muitas sinapses, em cada nível, executam muitos pequenos cálculos simultaneamente, em vez de um sequência laboriosa. Essa vantagem torna-se maior à medida que o número de neurônios aumenta a cada camada. Surpreendentemente, a velocidade de processamento é inteiramente independente, tanto do número de unidades envolvidas em cada camada quanto da complexidade da função que elas estão computando. Cada camada poderia ter quatro unidades ou cem milhões; sua configuração de pesos sinápticos poderia estar computando somas simples de um dígito ou equações diferenciais de segunda ordem. Não faria diferença. O tempo de computação seria exatamente o mesmo.

Em segundo lugar, o paralelismo massivo significa que o sistema é tolerante a falhas e funcionalmente persistente; a perda de algumas conexões, mesmo de um bocado, tem um efeito negligenciável sobre o caráter da transformação global realizada pela rede sobrevivente. Em terceiro lugar, um sistema paralelo armazena grandes quantidades de informação de forma distribuída, sendo que qualquer parte delas pode ser acessada em milissegundos. Essa informação é armazenada na configuração específica de forças de conexão sináptica, moldadas pelo aprendizado passado. As informações relevantes são "liberadas" à medida que o vetor de *input* passa – e é transformado – por aquela configuração de conexões.

O processamento paralelo não é ideal para todos os tipos de computação. Em tarefas que exigem apenas um pequeno vetor de *input*, mas muitos milhões de computações recursivas rapidamente reiteradas, o cérebro desempenha-se muito mal, enquanto as máquinas MS clássicas sobresaem-se. Essa classe de cálculos é muito grande e importante, de modo que as máquinas clássicas serão sempre úteis – de fato, vitais. Há, no entanto, uma classe igualmente grande de computações para as quais a arquitetura do cérebro é a tecnologia superior. Essas são as computações que tipicamente as criaturas vivas confrontam: reconhecer o perfil de

um predador em um ambiente ruidoso; lembrar instantaneamente como evitar seu olhar, fugir a sua abordagem ou se defender de seu ataque; distinguir alimentos de não alimentos e parceiros de não parceiros; navegar por um ambiente complexo e em constante mudança física/social; e assim por diante.

Finalmente, é importante notar que o sistema paralelo descrito não está manipulando símbolos de acordo com regras sensíveis a estruturas. Em vez de manipulação de símbolos, parece ser apenas uma das muitas habilidades cognitivas que uma rede pode ou não aprender a exibir. A manipulação de símbolos governada por regras não é o seu modo básico de funcionamento. O argumento de Searle dirige-se contra máquinas MS governadas por regras; transformadores de vetores do tipo que descrevemos não são, portanto, ameaçados por seu argumento do quarto chinês, mesmo se ele fosse correto – algo de que encontramos razão independente para duvidar.



SISTEMAS NERVOSOS abrangem muitas escalas de organização, desde moléculas **neurotransmissoras** (embaixo) até o cérebro e a medula espinhal como um todo. Níveis intermediários incluem neurônios individuais e circuitos feitos a partir de alguns neurônios, tais como aqueles que produzem orientação seletiva a um estímulo visual (no meio), e sistemas constituídos por circuitos, tais como aqueles que servem à linguagem (canto superior direito). Apenas a pesquisa pode decidir quão aproximadamente um sistema artificial tem de imitar a um sistema biológico para ser capaz de inteligência.

Searle está ciente dos processadores em paralelos, mas pensa que eles também serão desprovidos de conteúdo semântico real. Para ilustrar seu fracasso inevitável, ele esboça um segundo experimento de pensamento – a academia chinesa – que tem um ginásio cheio de pessoas organizadas em uma rede paralela. A partir daí o seu argumento procede como no quarto chinês.

Achamos essa segunda história bem menos persuasiva do que a primeira. Em primeiro lugar, é irrelevante que nenhuma unidade em seu sistema entenda chinês, já que o mesmo é verdadeiro para o sistema nervoso: nenhum neurônio no meu cérebro entende inglês, embora meu cérebro, como um todo, entenda. Em segundo lugar, Searle esquece de mencionar que sua simulação (usando uma pessoa por neurônio, além de uma criança veloz para cada conexão sináptica) exigirá ao menos 10^{14} pessoas, uma vez que o cérebro humano tem 10^{11} neurônios, cada um dos quais tem, em média, mais de 10^{3} conexões. Seu sistema exigirá toda a população humana de mais de 10.000 Terras. Um ginásio não começará a realizar uma simulação justa.

Por outro lado, se tal sistema fosse montado em uma escala cósmica adequada, com todos os seus caminhos fielmente modelados no caso humano, poderíamos, então, ter em nossas mãos um cérebro grande, lento, construído de modo estranho, mas ainda funcional. Nesse caso, a suposição padrão é com certeza que, dadas os *inputs* apropriados, ele pensaria, não que ele não poderia pensar. Não há garantia de que a sua atividade constituiria um pensamento real, porque a teoria de processamento de vetores, esboçada acima, pode não ser a teoria correta de como funciona o cérebro. Tampouco há qualquer garantia *a priori* de que não poderia estar pensando. Searle está, mais uma vez, confundindo os limites de sua imaginação atual (ou do leitor) com os limites da realidade objetiva.

O cérebro é um tipo de computador, embora a maioria de suas propriedades permaneça por ser descoberta. Caracterizar o cérebro como uma espécie de computador não é nem trivial nem frívolo. O cérebro computa funções, funções de grande complexidade, mas não à maneira da IA clássica. Quando se diz que cérebros são computadores, não deve inferir que são computadores em série, digitais, que são programados, que exibem a distinção entre hardware e software ou que têm de ser manipuladores de símbolos ou seguidores de regras. Os cérebros são computadores em um estilo radicalmente diferente.

Como o cérebro administra significados ainda é desconhecido, mas está claro que o problema vai além da utilização da linguagem e além de seres humanos. Um pequeno monte de terra fresca significa para uma pessoa, e também para coiotes, que um roedor (*gopher*) está nas redondezas; um eco com um certo carácter espectral significa, para um morcego, a presença de uma mariposa. Para desenvolver uma teoria do significado, mais deve ser conhecido sobre como os neurônios codificam e transformam sinais sensoriais, mais sobre a base neural da memória, aprendizado e emoção e também sobre a interação dessas capacidades e o sistema motor. Uma teoria do significado baseada nos neurônios pode exigir revisão das próprias intuições, que agora parecem tão seguras e que são tão exploradas livremente nos argumentos de Searle. Essas revisões são comuns na história da ciência.

Poderia a ciência construir uma inteligência artificial, explorando o que se sabe sobre o sistema nervoso? Não vemos nenhuma razão em princípio por que não. Searle parece concordar, embora qualifique sua afirmação dizendo que

"qualquer outro sistema capaz de causar mente teria que ter poderes causais (ao menos) equivalentes aos dos cérebros". Terminamos endereçando essa afirmação. Presumimos que Searle não está afirmando que uma mente artificial bem sucedida deve ter *todos* os poderes causais do cérebro, tais como o poder de sentir o cheiro ruim de algo apodrecendo, abrigar vírus lentos como kuru (causador da "febre de Kuru"), manchar-se de amarelo com peroxidase do rábano silvestre, e assim por diante. A exigência de paridade perfeita seria como exigir que um dispositivo voador artificial colocasse ovos.

Presumivelmente, ele quer apenas exigir de uma mente artificial todos os poderes causais relevantes, como ele diz, à inteligência consciente. Mas quais são eles, exatamente? Voltamos à querela sobre o que é e não é relevante. Esse é um lugar inteiramente razoável para um desentendimento, mas é uma questão empírica, a ser testada e comprovada. É porque tão pouco se sabe sobre o que faz parte do processo de cognição e semântica, que é prematuro ficar muito confiante sobre quais aspectos são essenciais. Searle sugere em vários lugares que todos os níveis, inclusive o bioquímico, têm de ser representados em qualquer máquina que seja uma candidata à inteligência artificial. Essa afirmação é quase certamente muito forte. Um cérebro artificial pode usar algo diferente de bioquímicos para alcançar os mesmos fins.

Essa possibilidade é ilustrada por uma pesquisa de Carver A. Mead no Instituto de Tecnologia da Califórnia. Mead e seus colegas usaram técnicas VLSI analógicas para construir uma retina artificial e uma cóclea artificial. (Em animais, a retina e a cóclea não são meros transdutores: ambos os sistemas incorporam uma rede de processamento complexa). Essas não são meras simulações em um mini-computador do tipo que Searle ridiculariza; são unidades de processamento de informação reais, que respondem, em tempo real, à luz real, no caso da retina artificial, e ao som real, no caso da cóclea artificial. Seus circuitos estão baseados na anatomia e fisiologia conhecidas da retina de gatos e da cóclea de corujas, seus *outputs* são dramaticamente semelhantes aos *outputs* conhecidos dos órgãos em questão.

Esses *chips* não usam substâncias neuroquímicas, portanto, neuroquímicos claramente não são necessários para obter os resultados evidentes. Naturalmente, não se pode dizer que a retina artificial vê alguma coisa, porque seu *output* não está ligado a um tálamo ou córtex artificial. Se o programa de Mead poderia ser mantido e se poderia construir um cérebro artificial inteiro, isso permanece por ser visto, mas não há nenhum indício agora que a ausência de bioquímicos torna-o quixotesco.

Nós, e Searle, rejeitamos o teste de Turing como uma condição suficiente para a inteligência consciente. Por um lado, nossas razões para fazê-lo são semelhantes: concordamos que também é muito importante como a função *input-output* é alcançada; é importante que os tipos certos de coisas estejam acontecendo dentro da máquina artificial. Por outro lado, as nossas razões são bem diferentes. Searle baseia sua posição em intuições de senso comum sobre a presença ou ausência de conteúdo semântico. Baseamos a nossa posição sobre falhas comportamentais específicas das máquinas MS clássicas e sobre as virtudes específicas de máquinas com arquiteturas mais parecidas com as do cérebro. Esses contrastes mostram que certas estratégias computacionais têm vantagens vastas e decisivas em relação a outras no que diz respeito a tarefas cognitivas típicas, vantagens que são empiricamente inevitáveis. Claramente, o cérebro está fazendo uso sistemático dessas vantagens computacionais. Mas esse não precisa ser o

único sistema físico capaz de fazê-lo. A inteligência artificial, em uma máquina não-biológica mas massivamente paralela, continua sendo uma perspectiva atraente e discernível.

* * *

Leituras adicionais:

DREYFUS, Hubert L. **What computers can't do: a critique of artificial reason**. New York: Harper & Row, 1972. [Disponível em português: **O que os computadores não podem fazer.** Rio de Janeiro: Casa do Livro Eldorado, 1975.]

CHURCHLAND, Paul M. **A neurocomputational perspective:** the nature of mind and the structure of science. Cambridge, MA.: The MIT Press, 1989.

CHUCHLAND, Patricia S. **Neurophilosophy:** toward a unified understanding of the mind/brain. Cambridge, MA.: The MIT Press, 1986.

DENNETT, Daniel C. Fast thinking. In: _____. **The intentional stance**. Cambridge, MA.: The MIT Press, 1987.

TURING, Alan M. Computing machinery and intelligence. In: **Mind**, v. 59, p. 433-460, 1950. [Disponível em português: Maquinário computacional e inteligência. In: Laurence BonJour e Ann Baker (Ed.), Filosofia: textos fundamentais comentados. Porto Alegre: Artmed, 2010.]