
Ensaio

Inteligência artificial e pensamento: redefinindo os parâmetros da questão primordial de Turing

Artificial intelligence and thinking: redefining parameters of Turing's primordial question

Diego Zilio✉

Programa de Pós-Graduação em Filosofia da Mente, Epistemologia e Lógica, Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP), Marília, São Paulo, Brasil

Resumo

A primeira parte do ensaio apresenta as idéias fundamentais de Turing que contribuíram para o desenvolvimento da ciência cognitiva. Assume-se que, embora tenha apresentado uma definição operacional de pensamento, Turing não consegue escapar do antropocentrismo, já que o teste baseado no jogo da imitação tem como parâmetro o ser humano. Consequentemente, o objetivo da ciência cognitiva influenciada por Turing passou a ser o de formalizar o pensamento humano. A possibilidade dessa tarefa é analisada na segunda parte do ensaio, na qual também são apresentadas as principais características do processo de raciocínio humano. O resultado dessa análise sugere que a formalização do pensamento humano em máquinas é uma tarefa muito difícil, senão impossível. Ressalta-se, todavia, que desse resultado não implica a negação da proposta de Turing. É preciso apenas redefinir os parâmetros de seu teste. © Cien. Cogn. 2009; Vol. 14 (1): 208-218.

Palavras-chave: Turing; ciência cognitiva; lógica; pensamento; raciocínio; modelos mentais.

Abstract

The first part of the essay presents Turing's fundamental ideas that contributed to development of cognitive science. Since the test based on imitation game has the human being as parameter, it is assumed that, despite his operational definition of thinking, Turing doesn't escape from anthropocentrism. Therefore, formalize human thinking has become the goal of cognitive science influenced by Turing. The possibility of this task is analyzed in the second part of the article, where are also presented the principal characteristics of reasoning in human. The result of this analysis suggests that the formalization of human thinking in machines is a very difficult task, if not an impossible one. However, this result doesn't imply the invalidation of Turing's proposal. Redefining parameters of his test is just what is needed. © Cien. Cogn. 2009; Vol. 14 (1): 208-218.

Keywords: Turing; cognitive science; logics; thinking; reasoning; mental models.

A proposta de Turing e a gênese da ciência cognitiva clássica

A primeira sentença do texto “Computing Machinery and Intelligence”, de Allan Turing (1950), é um convite à reflexão. O tema proposto é precisamente delimitado em sua questão primordial: “podem as máquinas pensar?” (p. 433). Com a questão posta, faltava apenas uma definição clara dos termos envolvidos. O que é “máquina”? O que é “pensamento”? Turing logo se deu conta de que esses são termos perigosos, pois há uma gama enorme de significados que os acompanha, o que acabaria por dificultar uma definição precisa. Com esse problema em mãos, o autor encontrou uma resposta no “jogo da imitação”. A versão “humana” do jogo seria assim: um participante faria perguntas a outros dois participantes sem poder vê-los e sem ter acesso direto às suas respostas, sendo essas apresentadas por um mediador. Ambos os participantes deveriam convencer com suas respostas serem mulheres. Dessa forma, o homem deveria persuadir quem fizesse as perguntas enquanto a mulher deveria provar que, de fato, ela é a mulher. O intuito do jogo seria descobrir qual dentre os dois participantes é o homem e, conseqüentemente, qual é a mulher. Turing, então, muda as regras do jogo colocando uma máquina no lugar de um desses participantes. Nessa nova situação, o objetivo seria descobrir qual, dentre os dois participantes, seria o ser humano e qual seria a máquina. Se conseguir agir por meio de suas respostas tal como um ser humano sem que o participante que faz as perguntas perceba, essa máquina seria considerada inteligente e, por conseqüência, um ser pensante. É importante ressaltar, todavia, que de maneira alguma Turing partiu do pressuposto de que os homens pensariam de maneira diferente se comparados às mulheres. Esse problema não era de seu interesse. O jogo da imitação entre homem e mulher foi apenas uma tática didática para se chegar ao verdadeiro teste de Turing; o teste em que uma máquina deveria passar-se por um ser humano para ser considerada um ser pensante.

De acordo com Dennett (1985/1998), Turing não estava interessado em criar uma máquina que pensasse da mesma forma que os seres humanos, nem queria estabelecer os parâmetros de validade das teorias sobre inteligência ou, de forma mais geral, sobre os processos cognitivos. Mas foi exatamente isso o que aconteceu. A idéia de Turing foi decisiva para o desenvolvimento da ciência cognitiva, especialmente em seu desdobramento na inteligência artificial simbólica. Há três motivos principais que justificam a influência de Turing. Primeiramente, seu teste estabeleceu, mesmo que indiretamente, a independência entre a estrutura material da máquina (o *hardware*) e sua função (o *software*). De certa forma, o autor propiciou uma nova forma de análise que, embora mecanicista, seria independente da matéria física (Pylyshyn, 1986). Afinal, a máquina não estaria visível ao participante, que só teria acesso às suas respostas. Essa divisão culminou na visão funcionalista da mente. Há dois princípios básicos do funcionalismo: (1) entender como a mente funciona implica conhecer os estados funcionais que a caracterizam; e (2) os estados funcionais podem ser realizados em qualquer configuração física. Para entender o que o isso significa, tomemos o exemplo clássico de Putnam (1967/1991), o precursor do funcionalismo no contexto da filosofia da mente. O autor, numa crítica incisiva à teoria da identidade mente-cérebro, indagou se seria correto afirmarmos que as dores são nada mais que estados cerebrais. Segundo o funcionalismo, a resposta é negativa. A dor seria um estado funcional resultante da relação entre os estímulos ambientais que modificam os estados corporais; entre outros estados funcionais (mentais); e entre as respostas comportamentais. Por exemplo, só é possível afirmar que, no cérebro humano, as dores estão relacionadas com os disparos de neurônios específicos por conta dos estados funcionais mentais e das respostas comportamentais que fazem parte dessa relação. Um alienígena poderia ter outra constituição física, mas, mesmo assim, possuir estados funcionais aos quais classificaríamos como dor. O funcionalismo

embasou a tese de que a mente, isto é, os estados funcionais, poderiam ser instanciados em qualquer configuração física. Putnam afirma, especificamente, que a máquina de Turing (da qual falaremos adiante), por ser universal, forneceu a estrutura básica para que isso fosse feito. Assim, o ponto de vista funcionalista é intrínseco à ciência cognitiva.

O segundo motivo que justifica a grande influência de Turing na ciência cognitiva está em sua definição clara e precisa sobre o que é o pensamento: pensar é processar informações a ponto de conseguir resolver problemas. Se a máquina conseguir com sucesso responder às perguntas feitas pelo participante do jogo de Turing, o que implica processar informações, a ponto de enganá-lo passando-se por um ser humano, ela seria considerada uma máquina pensante. Tanto a separação entre máquina e função quanto a definição operacional de pensamento propiciaram uma suposta emancipação do objeto de estudo. Pensar não seria mais uma característica dos seres humanos. Qualquer coisa que processasse informações, a ponto de conseguir com sucesso ser classificada como inteligente, pensaria. Em poucas palavras, Turing supostamente havia se livrado do antropocentrismo.

Finalmente, o terceiro motivo está no fato de que Turing também acabou por estabelecer o teste empírico em seu jogo da imitação. A máquina pensaria se conseguisse enganar o interlocutor. Com esse trabalho, Turing estabeleceu a agenda de pesquisa da ciência cognitiva, fazendo com que a computação, a matemática e a lógica trabalhassem juntas (Clark, 2001). Afinal, não há processamento de informações sem algoritmos. Uma definição precisa de algoritmo é apresentada por Knuth (1977): um conjunto de fórmulas, regras e parâmetros computáveis que possibilitam a produção de um conjunto específico de informações (*output*) quando na presença de um conjunto específico de informações (*input*). Uma máquina computa informações que chegam ao seu sistema de entrada (*input*). Essas informações são manipuladas de acordo com os algoritmos da máquina que, assim, apresenta uma resposta (*output*).

Turing (1950) também discorreu sobre o funcionamento de sua máquina. A *hipotética* máquina de Turing seria constituída por uma fita de dados de tamanho infinito, mas de estados finitos (“*finite state machine*”); por um processador de informações; e por um cabeçote capaz de ler, apagar e escrever informações na fita, além de poder movimentá-la. A máquina seria capaz de processar informações serialmente, com “memória” capaz de recordar qual a função do símbolo que está inscrito na fita e qual o estado da máquina no momento da leitura, podendo, assim, determinar a próxima ação (que é efetivada pelo cabeçote) e, conseqüentemente, o próximo estado finito da máquina (que está inscrito na fita). A universalidade da máquina de Turing consiste na possibilidade de imputar nela qualquer algoritmo, não havendo, assim, ao menos em princípio, limites para os tipos de processos que ela poderia instanciar. A máquina de Turing acabou por instituir o padrão de funcionamento de todas as máquinas digitais que conhecemos.

Assim temos uma definição clara dos termos envolvidos na questão primordial de Turing. A máquina a qual nos referimos é a máquina de Turing. Pensar é processar informações. E mais, a máquina pensa se conseguir resolver, por meio do processamento de informações, problemas a ela apresentados. É importante ressaltar a magnitude do teste de Turing. Não há regras para os problemas que poderão ser apresentados para a máquina em forma de questões. O interrogador poderá fazer perguntas a respeito de qualquer assunto, dentro de qualquer contexto, e da forma que quiser. Dennett (1985/1998) assevera que o teste de Turing é o mais difícil de todos. O autor cita como exemplo a máquina *PERRY*, que apresentava padrões de resposta típicos de pacientes psicóticos. *PERRY* enganou diversos psiquiatras que não conseguiam diferenciar suas respostas quando comparadas às de pacientes psicóticos humanos. Levando-se em conta o fato de que os psiquiatras faziam perguntas que eram tipicamente dirigidas aos pacientes psicóticos, é possível afirmar que *PERRY* passou no

teste de Turing? Para os críticos, não. Uma das críticas mais espirituosas, também citada por Dennett (1985/1998), foi a de Joseph Weizenbaum. O autor afirmou que assim como Kenneth Colby, criador de *PERRY*, ele também havia criado com sucesso uma máquina capaz de simular um problema mental humano. Entretanto, ressaltou o autor, por conta do baixo orçamento, ele teve que usar máquinas de escrever ao invés de computadores digitais. Sua máquina simulava respostas típicas da pacientes com autismo. A ironia do argumento é evidente e cumpre sua função. Colby fez um recorte minúsculo da realidade mental humana. Embora complexo, o repertório de respostas possíveis que se pode esperar de um paciente psicótico é finito. A força do teste de Turing, por sua vez, está em não estabelecer barreiras, isto é, não deve haver regras que limitem tanto o *modus operandi* da máquina quanto as perguntas (*i.e.*, os problemas) que a ela poderiam ser apresentadas.

Entretanto, há uma característica importantíssima no teste de Turing que ainda não foi aqui contemplada. A máquina pensaria se, e somente se, conseguisse com sucesso emular um ser humano, condição essa da qual se conclui que o parâmetro classificatório da máquina como pensante ou não-pensante é o próprio homem. Isto é, é o ser humano, enquanto ser pensante, quem profere o veredicto final. Turing, no final, não conseguiu se livrar inteiramente do antropocentrismo, já que o ser humano está no centro de seu teste. É o homem quem deve ser enganado. E o princípio da enganação seria a máquina ser comparada, sob a óptica do enganado, com um ser humano. Esse detalhe acabou por influenciar profundamente as pesquisas em ciência cognitiva, que, por um lado, atribuiu aos processos cognitivos humanos o status de fonte de dados primordial dos processos estudados, e, por outro lado, concentrou-se no desenvolvimento de máquinas na qual se pretendia simulá-los.

O ponto inicial desse processo de pesquisa estava na constatação de que o ser humano revela, por meio de suas ações, conhecimento sobre o mundo que o cerca. Consequentemente, uma boa estratégia para simular o pensamento em máquinas seria, então, formalizar esse conhecimento, isto é, transformar o conhecimento em algoritmos que possibilitem às máquinas agir tal como um ser humano, emitindo respostas específicas tipicamente “humanas” (*outputs*) na presença de problemas específicos tipicamente “humanos” (*inputs*). Hayes (1979/1990), por exemplo, propôs formalizar o conhecimento intuitivo humano a respeito do mundo físico. A “física ingênua”, nas palavras do autor, implicava formalizar o conhecimento humano a respeito dos objetos físicos, tratando de conceitos como “força”, “movimento” e “substância”, “liquidação”, entre outros, na construção de algoritmos que supostamente fariam com que a máquina interagisse com o mundo físico tal como os seres humanos. Mcdermott (1987/1990) sustenta que essa tentativa é um produto do “argumento logicista”, cujas premissas básicas seriam:

- (1) Um programa capaz de emular o pensamento humano deve possuir uma grande quantidade de conhecimento;
- (2) Esse conhecimento deve ser representado de alguma forma pelo programa;
- (3) Os programadores devem ter conhecimento a respeito do “conhecimento” que o programa precisa representar antes de desenvolver o programa em si;
- (4) A lógica e a matemática fornecem a notação necessária para representar o conhecimento nos programas;
- (5) Grande parte do conhecimento, senão todo, pode ser representado pela lógica dedutiva.

A ciência cognitiva influenciada por Turing assenta suas bases no argumento logicista e grande parte de seus problemas, como veremos adiante, decorreram desse fundamento. O argumento logicista apresenta uma forma de pensamento essencialmente lógico-formal e

mecanicista. O formalismo significa que os algoritmos são nada mais que regras lógicas. O mecanicismo, por sua vez, está relacionado com a questão da causalidade mecânica que norteia o funcionamento da máquina e é um princípio presente desde as pesquisas em cibernética que tiveram início durante a 2ª guerra mundial (Wisdom, 1951). É possível encontrar uma definição precisa sobre o assunto nos textos de Ashby (1947, 1962). Segundo o autor, o mecanicismo só faz sentido se relacionado ao modo de funcionamento da máquina, que é em si mecânico. Logo, o mecanicismo consiste na idéia de que sempre o estado finito presente da máquina determinará qual será o próximo estado finito. Se, de acordo com o algoritmo imputado na máquina, o estado finito “X” causar o estado finito “Y”, não haverá circunstância em que ocorrerá de outra forma. “Y” sempre seguirá de “X”. Assim, o mecanicismo implica uma forma de funcionamento, ou melhor, de causalidade, essencialmente determinística.

Resta-nos averiguar se o pensamento humano se enquadra nesse modelo. No âmbito das máquinas, Turing define pensamento como a capacidade para processar informações a ponto de conseguir resolver problemas. Por outro lado, a resolução de problemas se encontra, enquanto característica cognitiva, nos processos de raciocínio. Trata-se de definições correlatas. Logo, talvez seja viável buscar aproximações entre a proposta de Turing, no âmbito do pensamento das máquinas, e a proposta dos estudos da Psicologia Cognitiva sobre o processo de raciocínio.

Raciocínio e resolução de problemas

O raciocínio é definido, no contexto da Psicologia Cognitiva, como o processo de tomada de decisão (Giroto e Johnson-Laird, 1993; Johnson-Laird e Shafir, 1993; Leighton, 2004). Trata-se de uma característica do pensamento fundamentalmente dirigida para a resolução de problemas. As pesquisas nessa área buscam entender como se estrutura e qual a lógica de funcionamento do raciocínio. Deve-se compreender a passagem “lógica de funcionamento” no sentido literal, já que grande parte das pesquisas sobre o raciocínio abrange discussões da Lógica, sendo justamente essa característica que nos interessa neste ensaio. Nesse contexto, Johnson-Laird (2004b) desenvolveu a teoria dos modelos mentais. A idéia básica dessa teoria é que as pessoas utilizam o conhecimento geral que possuem do mundo para construir modelos mentais dos possíveis estados de coisas desse mundo. Ou seja, cria-se, pelos modelos mentais, um mundo hipotético onde a pessoa testa suas decisões antes de aplicá-las no mundo real. Há três características básicas dos modelos mentais. A primeira é que todos os modelos mentais representam uma possibilidade de estados de coisas do mundo. A segunda é que os modelos mentais são icônicos, isto é, suas constituições correspondem às constituições dos estados de coisas do mundo que representam. A terceira, definida como princípio da verdade, é que os modelos mentais representam apenas o que é verdadeiro em relação aos estados de coisas que os constituem. De acordo esse princípio, as pessoas normalmente não buscam delimitar o que é falso a partir de suas representações dos estados de coisas, mas, pelo contrário, buscam delimitar quais as possibilidades potencialmente verdadeiras para, assim, interagir com sucesso no mundo.

Os modelos mentais constituem a base fundamental de qualquer processo racional humano (Johnson-Laird, 1981, 1985, 2002, 2004a; Johnson-Laird *et al.*, 2004; Johnson-Laird e Yang, 2008). Entretanto, o raciocínio não é em si lógico no sentido formal do termo. Não há uma estrutura determinada para o raciocínio. As pesquisas de Johnson-Laird e de seus colaboradores mostram que cada pessoa constrói suas próprias estratégias para resolução de problemas. Tais estratégias são construídas ao longo da história de vida do sujeito e são mantidas de acordo com o sucesso em resolver problemas. A única constante, portanto, é a

existência desses modelos mentais e não a existência de uma estrutura lógica do pensamento. Essa constatação nos leva ao princípio da modulação, cuja idéia central é que o conhecimento prévio e as crenças de um sujeito modulam sua forma de raciocinar.

Há diversas pesquisas que confirmam a teoria dos modelos mentais. A estratégia básica implica constatar, dentre outras coisas, as similaridades e diferenças entre os processos de raciocínio dos sujeitos experimentais e as regras lógico-formais. Os resultados sugerem discrepâncias fundamentais entre esses dois âmbitos. Há diferenças significativas a respeito de como os sujeitos apreendem a noção de causalidade (Frosch e Johnson-Laird, 2006; Goldvarg e Johnson-Laird, 2001); a noção de lógica silogística (Bucciarelli e Johnson-Laird, 1999); a noção de necessidade e possibilidade da lógica modal (Evans *et al.*, 1999); e a noção de lógica clássica proposicional (Johnson-Laird, 2004a).

O intuito deste ensaio não é detalhar passo a passo os resultados dessas pesquisas. Sendo assim, um pequeno exemplo já nos fornece a base necessária para a argumentação conseguinte. O que todos os resultados apontam é que o processo de raciocínio, em pessoas que não possuem conhecimento de Lógica, quase nunca segue regras formais, mas é modulado pelo conhecimento e pelas crenças que os sujeitos possuem a respeito dos estados de coisas referidos pelos problemas. Dessa forma, por exemplo, tomemos a regra da disjunção, “A ou B. Não B. Então A.”, que no problema teria esta estrutura: João está em Bauru ou ao menos no estado de São Paulo. João não está no estado de São Paulo. Então, João está em Bauru. Embora formalmente correta, ninguém defenderia esta conclusão, pois se sabe que Bauru é uma cidade do estado de São Paulo e, se João não está nesse estado, tampouco está em Bauru. Assim, o conhecimento e as crenças, além de modularem a resolução de problemas, também afetam o próprio processo de raciocínio. O resultado dessa conclusão é que não há uma estrutura lógica determinada, e muito menos uma que tenha as regras lógico-formais como base, para os processos de tomada de decisão perante problemas.

As constatações de Johnson-Laird não são aceitas sem críticas pelos estudiosos do raciocínio (Bonati, 1994; O'Brien *et al.*, 1994). O'Brien (2004), por exemplo, apresenta uma teoria cuja premissa básica é justamente a presença de uma estrutura lógica determinada. Entretanto, ressalta o autor, não há razão alguma para crer que a lógica do raciocínio se equipare à lógica formal desenvolvida pelos logicistas. Isso significa que, mesmo que exista uma estrutura lógica determinada do raciocínio, é errado relacioná-la com as formas lógicas. Em poucas palavras, a lógica da mente não é a lógica formal. É evidente que a lógica formal decorre da lógica da mente, já que é resultado dos processos de pensamento dos seres humanos. Mas é errado pensar que, por conta disso, é possível cingir o processo de pensamento por meio de sistemas formais. Tal constatação é problemática ao “argumento logicista” da ciência cognitiva, pois sua proposta era justamente formalizar o pensamento.

Entretanto, uma das premissas do argumento logicista é que a máquina deveria representar o conhecimento que o homem possui para, assim, manipular as informações e chegar às respostas dos problemas apresentados a ela. O teste de Turing, por sua vez, exigiria que a máquina fizesse isso com tamanha acurácia a ponto de conseguir enganar o interlocutor. Voltemos-nos ao exemplo citado anteriormente. A intuição nos diz que, se a máquina tiver acesso à informação de que Bauru é uma cidade do estado de São Paulo, ela provavelmente chegará à mesma conclusão que as pessoas, subvertendo a regra lógica da disjunção. Mas suponha-se, então, que um dos algoritmos que norteiem o funcionamento da máquina seja propriamente a regra da disjunção. Se assim for, a máquina chegará à resposta contraditória de que João está Bauru mesmo não estando no estado de São Paulo. Mcdermott (1987/1990), antes um defensor ferrenho do argumento logicista, assevera que não importa a quantidade de informações imputadas na máquina porque, afinal, as inferências e deduções não decorrerão delas. Para esclarecer o que isso significa é interessante dividir conceitualmente o conteúdo e

o processo mental, em que o primeiro diz respeito ao conhecimento e as crenças das pessoas a respeito dos estados de coisas do mundo e o segundo é a forma como elas usam esse conhecimento e essas crenças para resolver problemas. Baseando-se nas constatações de Johnson-Laird e de seus colaboradores é factível sustentar que é impossível dividir tais âmbitos no processo de raciocínio. Ora, o conhecimento e as crenças modulam as estratégias de raciocínio das pessoas. Mcdermott (1987/1990), por outro lado, afirma que a ciência cognitiva apresenta uma notação essencialmente formal, o que significa que seu foco é apenas o processo. Nas máquinas, a estrutura formal do processo de pensamento precede o conteúdo informativo, manipulando-o conforme as regras lógicas implementadas. Assim, se a regra da disjunção, por exemplo, faz parte da “natureza” da máquina, esta a seguirá não importando os resultados que decorrerão disso. Para a máquina, João estará em Bauru, sendo irrelevante que tal resultado decorra do fato de que João não está no estado de São Paulo.

Alguns defensores do argumento logicista, na esperança de conseguir acabar com esse problema, desenvolveram máquinas capazes de modificar e de criar novos dados informativos (Mcdermott, 1987/1990). Mas o problema, mesmo assim, permanece. Mesmo se as máquinas fossem capazes de rever todo o seu “conhecimento” e colocar todas as suas “crenças” à prova, elas não poderiam modificar sua estrutura lógica de pensamento. O processo de revisão dos conteúdos e de reestruturação das estratégias de raciocínio, por sua vez, é uma característica essencial do pensamento humano e a impossibilidade de simulá-lo em máquinas sugere a conclusão pessimista de que as máquinas não poderão pensar tal como os homens (Johnson-Laird, 1981, 1985; Johnson-Laird e Yang, 2008). É possível criar máquinas peritas em diversos assuntos ou que sigam diversas regras lógicas, tal como *PERRY*, mas o teste de Turing exige algo mais. Como dissemos anteriormente, não deve haver barreiras ou regras que limitem tanto o modo de funcionamento da máquina quanto as perguntas que a ela serão feitas; assim como não há regras fixas que norteiam o processo de raciocínio humano.

Afinal, podem as máquinas pensar?

Na primeira parte deste ensaio, foram apresentadas as principais idéias de Turing a respeito da possibilidade de se criar uma máquina pensante. Ressaltou-se que há na ciência cognitiva, especialmente em seu desdobramento na inteligência artificial simbólica, o pressuposto do argumento logicista, segundo o qual seria possível formalizar o processo de pensamento, assim como o conhecimento humano, por meio das regras lógicas. A segunda parte, por sua vez, tratou da teoria dos modelos mentais de Johnson-Laird. A principal constatação extraída dessa teoria é que o raciocínio, enquanto processo de tomada de decisão perante problemas, fundamenta-se principalmente no conhecimento e nas crenças, que, por sua vez, modulam as estratégias para a resolução de problemas. A única constante nesse processo é a existência de modelos mentais que não são, em si, processos lógicos. Assim, a teoria dos modelos mentais sugere que há uma relação de síntese entre conteúdos (conhecimentos e crenças) e processos de raciocínio; relação essa denominada princípio da modulação. Nesse contexto, o termo “estratégias” para resolução de problemas cabe perfeitamente, já que não comprometeria o processo com alguma forma lógica fixa.

A questão mais importante, todavia, reside no fato de que não há normatividade nos processos de pensamento. Não há regras para a resolução de problemas, ou melhor, nas palavras de Johnson-Laird (2002), não há algoritmos. Podem existir similaridades entre sujeitos que estejam inseridos no mesmo contexto cultural, por exemplo, mas, mesmo assim, prevalecem as diferenças individuais nas estratégias de raciocínio (Lee e Johnson-Laird, 2006). Os sujeitos utilizam-se das estratégias que quiserem. O ajuste fino ocorrerá de acordo com o sucesso dessas estratégias na resolução de problemas. O fracasso, portanto, não implica

que o sujeito que falhou “não pensa” ou “não raciocina”. Tal liberdade, por sua vez, é impossível nas máquinas porque a natureza do processamento de informação é essencialmente mecanicista e normativa. Enfim, as máquinas podem rever seus conteúdos informativos e criar informações novas, mas nunca poderão mudar a sua própria “natureza” normativa lógico-formal, que, por sua vez, foram a elas atribuídas pelo homem enquanto ser pensante.

Essas constatações podem encorajar a defesa de uma distinção radical entre os processos lógico-formais e o pensamento. De fato, separar a Lógica do pensamento humano não é uma idéia nova. Frege foi o primeiro a estabelecer essa distinção com seu antipsicologismo (Oliveira, 1998). Para Frege, a Lógica não teria relações necessárias com os processos mentais, já que estes seriam subjetivos e privados enquanto ela seria objetiva e pública. Frege, então, propôs uma separação entre “pensamento” e “idéia”. A idéia seria o objeto de estudo da Psicologia, constituindo os eventos mentais privados e subjetivos. Já o pensamento seria um objeto abstrato, uma proposição. Assim definido, o pensamento seria acessível e público, podendo constituir o campo de estudo da Lógica. Entretanto, tal como sustenta Haack (1978/1998), o argumento de Frege não nos obriga a separar o pensamento dos processos mentais. Definir o pensamento como um objeto abstrato proposicional ainda deixa em aberto o problema de como o sujeito o apreende. Apresenta-se, então, um dilema: ao passo em que os processos de pensamento não são formalizáveis tal como a ciência cognitiva sustenta, tampouco é possível separar a Lógica da Psicologia.

Haack (1978/1998) afirma que há três posições básicas a respeito do debate entre Lógica e Psicologia. A primeira é o já citado antipsicologismo de Frege, que sugere uma distinção radical entre Lógica e Psicologia. A segunda posição, definida pela autora como descritivista, sustenta que a Lógica apresenta a descrição de como nós pensamos. Parece ser a posição sustentada pela ciência cognitiva. Entretanto, os dados apresentados anteriormente neste ensaio sugerem que o processo de pensamento, embora abarque a lógica formal, não é formalizável. Finalmente, a terceira posição é classificada como prescritiva. Como o termo sugere, a idéia fundamental seria que a Lógica apresenta as normas de como nós deveríamos pensar.

Talvez uma resposta possível ao problema da ciência cognitiva delineada há mais de cinquenta anos por Turing se encontre nessa última posição. Não haveria sentido em afirmar que a Lógica é prescritiva se os processos de pensamento fossem em si lógico-formais. Ora, a prescrição é um conjunto de regras, avisos e leis a respeito de como as coisas *devem* ser e não de como as coisas *são*. Assim, é plenamente possível sustentar que o pensamento humano não possui em si uma estrutura lógico-formal ao mesmo tempo em que se defende que a Lógica apresenta as formas corretas de como se deve pensar.

Nesse momento, é pertinente lembrar uma característica essencial das idéias de Turing: o antropocentrismo. No jogo da imitação o parâmetro responsável pela classificação das máquinas como pensantes ou não-pensantes é o ser humano. Nesse sentido, tendo o homem como referência, a pergunta central não seria apenas se as máquinas pensam, mas sim se as máquinas pensam tal como os seres humanos a ponto de enganá-los. Embora Turing tenha se esforçado para se livrar do antropocentrismo, seu teste acabou por aproximar a ciência cognitiva da Psicologia. A máquina deveria pensar como o ser humano. Só assim ela seria considerada um ser pensante. Todavia, os dados das pesquisas de Johnson-Laird e de seus colaboradores mostram que, neste caso, dificilmente (para não dizer impossível) haverá uma máquina que pensa.

Ao passo em que ofereceu uma definição operacional precisa de pensamento, o erro crucial de Turing foi não abandonar o antropocentrismo em seu teste. Esse problema acabou por direcionar a ciência cognitiva para um caminho problemático, em que havia a promessa

de que um dia máquinas pensariam tal como os seres humanos e de que era isso o que Turing queria provar em seu texto. O desenvolvimento histórico da ciência cognitiva foi, possivelmente, influenciado por isso. As dificuldades encontradas pelo modelo formal da inteligência artificial simbólica acabaram por fortalecer o programa conexionista das redes neurais artificiais. As dificuldades em simular os processos de pensamento do ser humano encontradas por esses modelos, por sua vez, levaram os cientistas cognitivos a crerem que o problema estava no fato de que as máquinas eram “paralíticas”, ou seja, que elas não possuíam corpo e nem interagiam com o ambiente; constatação que deu origem aos modelos da cognição incorporada e situada.

A despeito da importância de todos os modelos propostos pela ciência cognitiva – o intuito deste ensaio não é criticá-los ou apresentá-los de forma detalhada –, é preciso tomar cuidado para não abandonar o projeto de Turing por conta do erro antropocêntrico. Se a normatividade da Lógica for aceita, isto é, se, de fato, for consenso que a estrutura lógico-formal apresenta a maneira correta de como se deve pensar, é plenamente possível dar uma resposta positiva à questão primordial de Turing. A única condição necessária é que se abandone o parâmetro humano para classificar a máquina como pensante. Abandonando-se o teste de Turing, o que resta é uma definição de pensamento plenamente compatível com a de raciocínio humano, em que não há nenhuma sugestão sobre como *deve* ser o processo. Aceitando-se essas condições, é possível afirmar que as máquinas não pensam tal como os seres humanos, mas pensam tal como máquinas, isto é, de maneira mecanicista e seguindo normas lógico-formais.

Alguns autores, como Wittgenstein (citado por Shanker, 1987) e Piaget (1968), sustentam que mesmo nesses termos não seria possível afirmar que as máquinas pensam, pois o pensamento seria uma característica humana, não se tratando de um processo mecânico estruturado por regras lógicas. Caracterizá-lo dessa forma seria transgredir sua própria natureza. Todavia, devemos nos lembrar da definição de pensamento proposta por Turing: pensamento é o processo pelo qual se resolve problemas manipulando-se informações. Essa definição está de acordo com a de raciocínio apresentada pela Psicologia Cognitiva. Assim, algo que pensa é algo que resolve problemas, não importando se o processo pelo qual se chega às respostas seja lógico-formal e mecanicista ou baseado em modelos mentais e estratégias. Dessa forma, há tanto homens quanto máquinas plenamente capazes de resolver problemas.


Concluindo, já que este parece ser o único âmbito, dentre os citados por Haack (1978/1998), em que é possível unir sem problemas a Psicologia e a Lógica, talvez a questão mais importante que nos resta, afinal, seja essencialmente normativa: deveriam os seres humanos pensar tal como as máquinas?

Referências bibliográficas

- Ashby, W.R. (1947). The nervous system as physical machine: with special reference to the origin of adaptive behavior. *Mind*, 56 (221), 44-59.
- Ashby, W.R. (1962). Principles of the self-organizing system. Em: Foersterand, H.V. e Zopf, G.W. (Eds.), *Principles of self-organization: transactions of the University of Illinois symposium* (pp. 255-278). London: Pergamon Press.
- Bonatti, L. (1994). Prepositional reasoning by model? *Psychological Rev.*, 101, 725-733.
- Bucciarelli, M. e Johnson-Laird, P.N. (1999). Strategies in syllogistic reasoning. *Cognitive Sci.*, 23 (3), 247-303.
- Clark, A. (2001). *Mindware: an introduction to the philosophy of cognitive science*. New York: Oxford University Press.

- Dennett, D. (1998). Can machines think? Em: Dennett, D., *Brainchildren: essays on designing minds* (pp. 3-29). Cambridge: The MIT Press. (Original publicado em 1985).
- Evans, J.; Handley, S.; Harper, C. e Johnson-Laird, P.N. (1999). Reasoning about necessity and possibility: a test of the mental model theory of deduction. *J. Exp. Psychol. Learn. Memory Cogn.*, 25 (6), 1495-1513.
- Frosch, C.A. e Johnson-Laird, P.N. (2006). The revision of beliefs about causes and enabling conditions. Em: Sun, R. (Ed.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 1329-1333). New Jersey: Erlbaum.
- Giroto, P.L. e Johnson-Laird, P.N. (1993). Focussing in reasoning and decision making. *Cognition*, 49, 37-66.
- Goldvarg, E. e Johnson-Laird, P.N. (2001). Naive causality: a mental model theory of causal meaning and reasoning. *Cognitive Sc.*, 25, 565-610.
- Haack, S. (1998). *Filosofia das Lógicas* (Mortari, C.A. e Dutra, L. Trads.). São Paulo: Editora Unesp. (Original publicado em 1978).
- Hayes, P.J. (1990). The naïve physics manifesto. Em: Boden, M. (Ed.), *The Philosophy of Artificial Intelligence* (pp. 171-205). New York: Oxford University Press. (Original publicado em 1979).
- Johnson-Laird, P.N. (1981). Cognition, computers, and mental models. *Cognition*, 10, 139-143.
- Johnson-Laird, P.N. (1985). Human and computer reasoning. *Trends Neurosci.*, 8, 54-57.
- Johnson-Laird, P.N. (2002). Logic and reasoning. Em: Ramachandran, V.S. (Ed.), *Encyclopedia of the human brain* (Vol. 2, pp. 703-716). San Diego: Academic Press.
- Johnson-Laird, P.N. (2004a). Mental models and reasoning. Em: Leighton, J.P. e Sternberg, R.J. (Eds.), *The Nature Of Reasoning* (pp. 169-204). New York: Cambridge University Press.
- Johnson-Laird, P.N. (2004b). The history of mental models. Em: Manktelow, K. e Chung, M. (Eds.), *Psychology of Reasoning: Theoretical and Historical Perspectives* (pp. 179-212). New York: Psychology Press.
- Johnson-Laird, P.N.; Legrenzi, P. e Giroto, V. (2004). How we detect logical inconsistencies. *Curr. Directions Psychological Sc.*, 13 (2), 41-45.
- Johnson-Laird, P.N. e Shafir, E. (1993). The interaction between reasoning and decision making: An introduction. *Cognition*, 49, 1-9.
- Johnson-Laird, P.N. e Yang, Y. (2008). Mental logic, mental models, and computer simulations of human reasoning. Em: Sun, R. (Ed.), *Cambridge Handbook of Computational Cognitive Modeling* (pp. 339-358). Cambridge: Cambridge University Press.
- Knuth, D. (1977). Algorithms. *Scientific American*, 236 (4), 63-80.
- Lee, N.Y.L. e Johnson-Laird, P.N. (2006). Are there cross-cultural differences in reasoning? Em: Sun, R. (Ed.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 459-464). New Jersey: Erlbaum.
- Leighton, J.P. (2004). Defining and describing reason. Em: Leighton, J.P. e Sternberg, R.J. (Eds.), *The Nature of Reasoning* (pp. 3-11). New York: Cambridge University Press.
- Mcdermott, D. (1990). A critique of pure reason. Em: Boden, M. (Ed.), *The Philosophy of Artificial Intelligence* (pp. 206-230). New York: Oxford University Press. (Original publicado em 1987).
- O'Brien, D.P. (2004). Mental-logic theory: What it proposes, and reasons to take this proposal seriously. Em: Leighton, J.P. e Sternberg, R.J. (Eds.), *The Nature of Reasoning* (pp. 205-233). New York: Cambridge University Press.
- O'Brien, D.P.; Braine, M.D.S. e Yang, Y. (1994). Propositional reasoning by mental models? Simple to refute in principle and in practice. *Psychological Rev.*, 101, 711-724.

- Oliveira, M.B. (1998). Lógica e ciência cognitiva: o antipsicologismo de Frege. Em: Oliveira, M.B. (Ed.). *Da Ciência Cognitiva à Dialética* (pp. 27-50). São Paulo: Discurso Editorial.
- Piaget, J. (1968). A explicação em psicologia e o paralelismo psicofísico. Em: Piaget, J.; Fraise, P. e Reuchlin, M. (Eds.), *Tratado de Psicologia Experimental: Volume 1* (pp. 121-152). Rio de Janeiro: Editora Forense.
- Putnam, H. (1991). The nature of mental states. Em: Rosenthal, D.M. (Ed.), *The Nature of Mind* (pp. 197-203). New York: Oxford University Press. (Original publicado em 1967).
- Pylyshyn, Z. (1986). *Computation and cognition*. Cambridge: The MIT Press.
- Shanker, S.G. (1987). Wittgenstein versus Turing on the nature of Church's thesis. *Notre Dame J. Formal Logic*, 28 (4), 615-649.
- Turing, A. (1950). Computing machinery and intelligence. *Mind*, 59 (236), 423-460.
- Wisdom, J.O. (1951). The hypothesis of cybernetics. *Br. J. Philosophy Sci.*, 2 (5), 1-24.

 - **D. Zilio** é Psicólogo (UNESP - campus de Bauru) e Pós-graduando do Programa de Pós-graduação em Filosofia da Mente, Epistemologia e Lógica (Faculdade de Filosofia e Ciências, UNESP – Marília). Endereço para correspondência: Rua Olavo Bilac, 15-5, Bairro Bela Vista, Bauru, SP 17060-454. *E-mail* para correspondência: diego.zilio@terra.com.br.