
Artigo Científico

A crise do computacionalismo: por uma nova metáfora computacional

The crisis of the computationalism: toward a new computational metaphor

Gustavo Arja Castañón[✉]

Universidade Estácio de Sá, Rio de Janeiro, Brasil e Universidade Católica de Petrópolis,
Petrópolis, Rio de Janeiro, Brasil

Resumo

A teoria do computacionalismo pode ser resumida em três alegações centrais, a de que “o cérebro é um computador digital”, que “a mente é um programa computacional” e que “as operações do cérebro podem ser simuladas em um computador digital”. Aqui serão expostas as refutações de John Searle, Hubert Dreyfus, Peter Jackson, Thomas Nagel e Joseph Rychlak a estas alegações. Em virtude da profundidade destas críticas e de seus problemas, conclui-se que, como concebido originalmente pelo Funcionalismo de Hilary Putnam, o computacionalismo está superado. Apesar disto, será aqui exposta tese original de que a força da metáfora do cérebro como computador ainda pode ser útil para o auxílio da compreensão humana sobre a mente e os aspectos irredutíveis da consciência, se substituímos os dois níveis de análise cérebro/hardware e mente/software pelos três níveis cérebro/hardware, processos cognitivos/software e consciência/usuário. © Ciências & Cognição 2006; Vol. 09: 27-41.

Palavras-chave: metáfora computacional; computacionalismo; consciência; inteligência artificial; filosofia da mente.

Abstract

The theory of computationalism can be summarized in three central allegations, that "the brain is a digital computer", that "the mind is a computational program" and that "the operations of the brain can be simulated in a digital computer". Here, will be exposed the John Searle, Hubert Dreyfus, Peter Jackson, Thomas Nagel and Joseph Rychlak' refutations to these allegations. Because of the depth of these critics and their problems, was concluded that, as originally conceived by Hilary Putnam's Functionalism, the computationalism is overcome. In spite of this, here it will be exposed original theory that the power of the metaphor of the brain as computer can still be useful for the aid of human understanding on the mind and unyielding aspects of the consciousness, if we substitute the two levels of analysis brain/hardware and mind/software for the three levels brain/hardware, cognitive processes/software and consciousness/user. © Ciências & Cognição 2006; Vol. 09: 27-41.

Keywords: *computational metaphor; computationalism; consciousness; artificial intelligence; philosophy of the mind.*

✉ - **G.A. Castañón** é graduado em Psicologia (UERJ) e em Filosofia (UFRJ), mestre em Psicologia Social (UERJ) e doutor em Psicologia (UFRJ). Atualmente ministra cursos na graduação em Psicologia da Universidade Estácio de Sá, e cursa o Mestrado em Lógica e Metafísica na UFRJ, tendo se dedicado nos últimos oito anos à investigações de Filosofia da Psicologia. *E-mail* para correspondência: gustavocastanon@hotmail.com.

Em artigo intitulado “John Searle e o Cognitivismo”, Castañon (2006) introduz o problema do computacionalismo hoje, assim como as críticas de Searle às teses centrais desta abordagem da mente humana. Aceitei a definição de ‘computacionalismo’ como a crença de que os processos mentais consistem essencialmente em manipulação simbólica lógica, ou seja, computação. É fundamental para uma adequada caracterização desta abordagem da mente recorrer a Jerry Fodor e a sua formulação da *computational theory of mind* (CTM). Em seu clássico “*Language of Thought*”, de 1975, Fodor apresenta uma versão computacionista onde nossos estados mentais são estados simbólicos associados numa seqüência de outros símbolos dentro de algum tipo de linguagem do pensamento (como a linguagem de máquina), e processos mentais são transformações lógicas destas cadeias de símbolos. A mais importante das heranças legadas pelo computacionalismo é a chamada “metáfora computacional” (Neisser, 1967), apropriada pelo cognitivismo menos pelas possibilidades futuras de simulação de processos cognitivos em máquinas do que pela clareza conceitual que a distinção entre hardware e software permitiu à teorização sobre a mente e suas relações com o cérebro.

Para a “metáfora computacional”, o nível de análise do cérebro era o nível do hardware, da *máquina*; o nível de análise da mente, era o nível do software, do *programa*. Esta idéia introduzida por Hilary Putnam (1961) foi uma das mais influentes do século XX, sedimentando um programa de pesquisa que logo seria conhecido pelo termo *inteligência artificial* (IA). Seu objetivo era projetar um computador e um programa que operassem de forma idêntica a um ser humano, o que nos levaria a compreender nossos próprios processos cognitivos. Era o nascimento da tese da IA forte.

Searle (1992) define o computacionalismo como sustentado em três alegações. A primeira destas alegações seria a de que “o cérebro é um computador digital”. A segunda, que “a mente é um programa computacional”. A terceira, que “as operações do cérebro podem ser simuladas em um computador

digital”. Como visto por Castañon (2006), estas três alegações são refutadas, em argumentos que devem ser resumidos brevemente aqui.

Searle e a crítica ao computacionalismo

Vamos lembrar como Searle refuta as três alegações do computacionalismo, começando pela segunda, de que “a mente é um programa computacional”. Neste argumento, ele evidencia que a dimensão sintática não é suficiente para explicar o que faz a mente. Se pensar é manipular símbolos através de regras puramente formais e sintáticas, então a semântica, o significado das representações, não tem lugar na explicação psicológica. Com sua famosa metáfora do quarto chinês, Searle (1984) evidencia a inviabilidade de tal concepção.

Argumenta Searle (1984) que se um programador que não sabe absolutamente nada de chinês fosse trancado num quarto de hotel na China com livros de regras de transformação de ideogramas chineses em outros ideogramas chineses, e recebesse todo dia pela manhã uma pergunta em chinês sobre a interpretação de um texto para passar o dia aplicando as regras nela, poderia conseguir chegar ao fim do dia a outros ideogramas que seriam respostas adequadas sobre o texto. Sua tarefa, assim como a do computador do teste de Turing, é aplicar uma série de operações formais (um programa) a símbolos que recebe como input (ideogramas-perguntas sobre uma história), produzindo ao final da aplicação outros símbolos como output (ideogramas-respostas adequadas a estas perguntas). Um avaliador chinês que analisa estas respostas, sabendo que foram resultado de um programa (expresso no livro de regras), pode chegar a afirmar que o suposto “computador” que o rodou compreendeu perfeitamente o texto. Mas o problema é que o programa foi “rodado” por um ser humano, que não leu o texto nem entendeu o significado de símbolo nenhum. Mas apesar de não ter compreendido o significado de qualquer frase ou símbolo, o programador processou informação de acordo com regras lógicas. Ou seja, aqui, o ser

humano fez o que um computador faz, e não compreendeu nada. Logo, obviamente, quando um computador executar o programa tão pouco estará compreendendo qualquer coisa. O que fica óbvio aqui, é que há uma dimensão do pensamento humano que não é reproduzível pelo computador: a semântica. O que fica evidente é que o que a mente humana faz, não é o que um computador faz. Um computador processa informação, para que nós evoquemos significados, que atribuímos às informações que o computador nos transmite.

Assim, como alegado anteriormente (Castañon, 2006), se temos um programa de computador onde nós colocamos uma série de características anatômicas sobre um animal e ele nos dá o nome de um animal com estas características, ele chegou a esta resposta através da aplicação de uma série de regras formais. O computador não sabe o que é uma característica anatômica, o que é um animal, ou um touro, ou um leão. Assim, saber a que estes símbolos se referem, qual é o significado do input e do output, não ajuda na tarefa de explicar como o computador chega à suas respostas.

Searle (1992), na obra “*The Rediscovery of the Mind*” (que pode vir a se tornar marca do fim do projeto da inteligência artificial e da CTM clássica), aprofunda sua crítica ao computacionalismo com dois argumentos contra as alegações primeira e terceira desta abordagem. A terceira das teses centrais do computacionalismo é segundo Searle (1992) de que “as operações do cérebro podem ser simuladas em um computador digital”. A esta afirmação Searle responde que *sim*, as operações do cérebro podem ser simuladas em um computador digital. As do cérebro, as das moléculas de um composto químico, das condições meteorológicas, do crescimento de uma planta e de tudo o que obedecer a padrões em todo o universo. Mas assim como ao simular o comportamento de um furacão nós não produzimos um furacão nem todas as suas propriedades, ao simular o comportamento de um cérebro nós não produzimos a consciência e suas propriedades emergentes.

Quanto à primeira alegação, a central, a tese fundamental do computacionalismo, de que “o cérebro é um computador digital”, nos deparamos com uma refutação surpreendente. Como em relação à alegação três, ele responde *sim*. *Um cérebro é um computador digital porque*, em última análise, seguindo as definições dadas por Turing, *tudo é um computador digital*. O que Searle está afirmando, é que o princípio de realizabilidade universal (um programa pode ser executado por qualquer coisa organizada para reagir digitalmente: de máquinas hidráulicas a um grupo de pessoas no painel do ursinho Misha da olimpíada de Moscow) de Putnam (1961) é válido porque a “sintaxe” não é algo físico como gravidade ou massa, e se encontra somente “nos olhos do observador”. Com esta inversão explicativa Searle (1992) demonstra que, não só as crenças básicas do computacionalismo são falsas, como possivelmente não tem um sentido muito claro. Qualquer coisa pode ser interpretada como um estado computacional, *menos a consciência*. A consciência é real, o programa computacional está nos olhos de quem vê (só existe para a consciência). Assim, a intencionalidade (a capacidade de algo se referir, direcionar ou significar algo além dele) não pode ser explicada muito menos identificada com computação.

A conclusão de Searle (1992) que mais interessa para a tese que será apresentada neste artigo, é a de que o computacionalismo é uma versão disfarçada da falácia do homúnculo. Esta falácia é apresentada resumidamente no último item deste artigo. Para Searle o CTM é uma nova tese do homúnculo porque trata o cérebro como se houvesse algum agente dentro dele usando-o para computar junto com ele. Quando compramos um computador numa loja, instalamos programas nele e o utilizamos para certas finalidades, não precisamos nos preocupar com o problema do homúnculo, porque *o homúnculo aqui somos nós*. Os sinais elétricos transformados em luz na tela do computador se tornam representações *para nós*. A “intencionalidade” do computador é a nossa, porque ela é somente derivada. Nós é

que associamos uma palavra a um significado, nós é que decidimos rodar este ou aquele programa. Mas o que significa dizer então que um computador processa informação? Nenhum computador processa completamente informação, diz Searle (1992), e assim nenhum cérebro também o faz. Computadores nos ajudam a transmitir informação e a processá-la, eles não fazem a pior parte do serviço. Um agente externo (programador) codifica alguma informação cujo significado já está previamente acordado e estabelecido, de uma forma que possa ser processada pelo hardware. O que o computador faz é transformar novamente, através de uma série de estágios, os sinais elétricos em signos (na tela, numa folha impressa) nos quais a informação está codificada, para que um agente externo (uma pessoa para a qual a informação tem um significado previamente estabelecido) possa interpretá-la *tanto sintática quanto semanticamente*, uma vez que o hardware não tem nenhuma sintática ou semântica intrínseca: “*It’s all in the eye of the beholder*” (Searle, 1992: 223).

Dreyfus e o que os computadores não podem fazer

O filósofo Hubert Dreyfus é uma figura ímpar no panorama do pensamento contemporâneo. Ele é, ao mesmo tempo, americano, fenomenólogo, e um dos mais importantes nomes do campo da inteligência artificial. Mais precisamente, é o mais famoso adversário da tese da IA forte. Em 1972, depois de alguns artigos que semearam a polêmica no âmbito da Inteligência Artificial, ele reuniu suas observações em um livro que se tornou um marco da Filosofia da Mente: “*What Computers Can’t Do*”. Neste ele apresenta algumas teses básicas que a longo prazo se revelariam insuperáveis pela IA, e que ainda hoje pautam a maioria das tentativas de expansão do campo.

A crítica de Dreyfus (1972) que mais atinge as pretensões do computacionalismo é a que advoga a impossibilidade de que seres humanos produzam inteligência usando somente fatos e regras. Para Dreyfus, as

dificuldades inerentes ao modelo de mente do processamento de informação são que este não possui senso de relevância em relação à informação a ser usada ou coletada, uma vez que representações simbólicas são atomistas e nosso senso de relevância é holista, no estilo gestáltico da palavra. Assim, Dreyfus (1972) predisse que o sonho de Turing estava condenado: um computador não seria capaz de responder com adequação (simulando entendimento) sequer como uma criança de quatro anos pode responder ao ouvir uma história infantil. Isto se daria em virtude do que seria um velho sonho racionalista, que ele erroneamente atribui a Leibniz, de que todo pensamento não passaria de computação lógica. Leibniz previu que o pensamento lógico, mais precisamente o raciocínio silogístico, poderia ser formalizado e reproduzido em máquinas que ele denominou “máquinas de julgar”, mas a visão de racionalidade leibniziana vai muito além do cálculo proposicional, ela é virtualmente infinita, como é o conteúdo de idéias contidas na mônada.

Dreyfus (1972) alega que o tipo de processo de resolução de problemas possuído por um expert de alguma área é diferente do tipo de procedimento seqüencial executado por um computador digital. A utopia da IA está baseada na crença de que todo nosso conhecimento sobre o mundo pode ser representado na forma proposicional, como um sistema de crenças implícitas. Assim, a tarefa hercúlea da IA seria a de derivar o senso comum de uma criança de quatro anos de uma gigantesca base de dados proposicionais (de crenças expressas na forma de sentenças lingüísticas comuns, porém, formalizadas) e criar as regras para computar este enorme volume de dados. O problema é que o conhecimento necessário para responder de forma a simular o entendimento da mais banal passagem de uma história infantil em inglês, requer um conjunto de conhecimentos formalizados do contexto, do falante e do mundo que está muito além da capacidade dos programas de computador (e dos computadores) conhecidos. E isto vale ainda hoje.

Para exemplificar o problema envolvido, vamos avaliar a seguinte seqüência de frases: “João viu uma bola na janela. Ele a quer.” Nada poderia ser mais banal e simples, e uma criança de três anos está perfeitamente apta a compreender a seqüência. Mas um computador tem extrema dificuldade em efetuar respostas que simulem entendimento desta frase. Pois a que se referiria o “a”? À bola ou à janela? E se mudássemos a segunda frase para “Ele a chutou”, ou “Ele a quebrou”, melhoraria a situação do computador? Não. Para Dreyfus isto se dá porque nossas habilidades de contexto são fruto da nossa capacidade de nos colocarmos no lugar dos outros e imaginar a nós mesmos na situação em questão. Não se trata de uma busca de fatos expressos proposicionalmente, tais como “crianças querem bolas e não janelas” ou “bolas, exceto se forem de natal, não quebram” e “janelas quebram”. No último caso, faltaria contexto para saber se o período é de natal e, sendo, se a bola era de natal, e assim por diante. Assim, sentencia Dreyfus (1972), nós precisamos ser capazes de nos imaginar sentindo e fazendo coisas para organizar o conhecimento que precisamos para compreender as sentenças típicas da nossa linguagem cotidiana.

Dreyfus (1993), aprofundando o argumento, avalia a tentativa da Inteligência Artificial de simular de outra maneira o entendimento de sentenças ordinárias. Supondo que tivéssemos um computador com milhões de fatos organizados sem propósito particular nenhum, como este poderia ser capaz de compreender uma sentença proferida numa situação específica? Isto é bem pior do que considerar o problema de compreender uma história infantil completa em si mesma.

De fato, hoje se tornou explícita no campo da IA esta evidente diferença entre o processamento computacional e o humano. Porque quanto mais proposições são acumuladas num sistema sobre um particular estado de coisas, situação ou tarefa, mais tempo demora para o sistema processar o que seria a informação relevante. Como sabemos, com a inteligência humana se dá o oposto: um expert vê instantaneamente não só a

informação relevante, como também o problema e a solução, enquanto o iniciante que aplica o livro de regras e truques “faça você mesmo” demora um tempo considerável para identificar a mais básica informação relevante. O que é ainda pior, como nos revelam os estudos cognitivos da memória, quanto mais sabemos mais fácil e rapidamente adquirimos novas informações, e as recuperamos também com mais facilidade. O contrário ocorre com a memória computacional: quanto mais informação, mais tempo de processamento. Algo portanto, por mais rápido que se torne um dia o processamento num supercomputador, está fundamentalmente errado com a concepção digital de inteligência humana. Nós construímos nossa memória de forma fundamentalmente diferente da simbólica representacional que os computacionalistas advogam.

Outras questões cruciais foram levantadas por Dreyfus (1972). Uma é a incapacidade de computadores apresentarem comportamento simulando a compreensão de analogias e menos ainda as utilizando. Imagine um computador tentando vencer o teste de Turing respondendo a seguinte pergunta: “Como é possível que Maria não tenha destruído completamente as pretensões da empresa onde foi explorada a vida toda, quando esta lhe propôs um acordo no meio daquela batalha judicial?” Esta frase comum, envolve uma série de analogias comuns: destruído, explorada, vida toda, batalha. Estas não são palavras que estão sendo usadas no seu contexto ou com seu significado ordinário. Um computador digital tem profunda incapacidade de responder adequadamente a estes problemas. Como aponta Searle (2000), isto se dá porque analogias são formas de pensamento totalmente não representacionais.

Outro grupo de problemas abordados por Dreyfus (1972) é o das dificuldades que computadores apresentam para reconhecer padrões. Isto se dá para ele em virtude da forma elementarista como computadores digitais tratam seus dados. Esta consiste em mais uma distância intransponível, pois

ontológica: a mente humana é holística, e a percepção funciona de maneira gestáltica, reconhecendo padrões e estruturas, não combinando elementos. Esta diferença básica entre a mente holística e analógica humana e a “mente” digital computacional é responsável por fracassos da IA em jogar xadrez (Dreyfus escreve antes do *Deep Blue...*), resolver problemas, reconhecer similaridades, reconhecer objetos em movimento, reconhecer faces e assim por diante.

Assim, por estes e outros problemas, Dreyfus (1972) conclui que se considerarmos a evidência descritiva fenomenológica sem preconceitos filosóficos, somos obrigados a concluir que existem capacidades humanas não programáveis em todas as formas de comportamento inteligente. Assim prevê que, uma vez que a Inteligência Artificial se trata de um problema empírico, não haveria grandes progressos nas áreas em questão. Vinte anos depois de feita esta profecia, Dreyfus (1993) relançou sua obra já clássica acrescida de uma revisão do trabalho em Inteligência Artificial nos anos que se seguiram, como o conexionismo e a continuação do programa de pesquisa da abordagem computacional da mente. Na apresentação de seu livro “*What Computers Still Can’t Do*” ele afirma que a nova edição de sua obra não marcava somente a mudança de título ou de editor (a editora da versão revista agora era a do MIT, berço da Inteligência Artificial), mas sim uma mudança de status: há vinte anos, ela representava uma posição controversa, agora, ela representava a nova posição padrão. De fato, as coisas se passaram como ele havia previsto. Das quatro categorias de atividades inteligentes, ele previu que duas seriam plenamente computáveis, uma insatisfatoriamente computável e a quarta de forma alguma computável.

Essas “categorias de atividade inteligente” eram a associacionista, a formal-simples, a formal-complexa e a não-formal. A atividade associacionista seria a caracterizada pela irrelevância do sentido e da situação. Seria inata ou aprendida por repetição (e.g.

jogo da memória, tradução palavra-a-palavra, condicionamento clássico). O tipo de programa que poderia reproduzir este comportamento seria o de árvore de decisão ou de busca em lista. A atividade formal-simples é aquela onde os significados já estão completamente explícitos e são independentes da situação, sendo aprendida por regras (ex: prova de teoremas usando procedimentos de lógica clássica, jogos computáveis, problemas de combinatória). Esta espécie de inteligência Dreyfus identifica com o *esprit de géométrie* de Pascal, e o tipo de programa que pode simular a atividade inteligente nestes âmbitos é o algorítmico. A atividade formal-complexa é segundo Dreyfus semelhante ao segundo tipo de programa, mas na prática é dependente de estados internos em situações específicas, só sendo adquirida pela prática na aplicação das regras (ex: jogos complexos como xadrez, problemas complexos de combinatória que envolvam decisões de planejamento, teoremas lógicos que envolvam intuições em passos da demonstração, reconhecimentos de regularidades em situações normais). Aqui o tipo de programa teria que ser de busca heurística, ou seja, capaz de discriminar alguma relevância na enorme massa de dados a computar. Dreyfus prevê em 1972 que programas deste tipo eram possíveis mas seriam formas de executar as tarefas muito ineficientes e abaixo da expertise humana.

Em grande parte pela influência dos argumentos de Dreyfus é que tamanha expectativa se criou em torno do segundo desafio em 1997 entre o supercomputador *Deep Blue* e Gary Kasparov, o maior mestre de xadrez de todos os tempos. Depois de uma vitória dramática que decidiu a série para o *Deep Blue* na última partida, Kasparov declarou que aquele era o fim da espécie humana. Não era. Apesar do avanço dos programas heurísticos e da capacidade computacional dos supercomputadores, hoje é aceito generalizadamente que o tipo de processo heurístico realizado por humanos é muito diverso daquele realizado por supercomputadores ao simular a expertise humana nessas atividades. E de fato, a previsão de

Dreyfus quanto à impenetrabilidade do computador no quarto campo de atividade inteligente, a não-formal, se revelou até aqui totalmente acertada.

Atividade inteligente não-formal é para Dreyfus (1972) aquela dependente de significados não-explícitos e de contextos, e quando aprendida, só o é de maneira intuitiva através de metáforas ou exemplos perspicazes de comportamento (jogo de imagem e ação, insights sobre problemas de estrutura aberta, tradução da linguagem natural em contexto de uso – com o caso extremo da poesia – e reconhecimento de padrões distorcidos). Este tipo de inteligência Dreyfus identifica com o *esprit de finesse* de Pascal, e decreta que não há, nem nunca haverá, ao menos com a atual tecnologia de computação digital, qualquer tipo de programa que a possa simular.

Concluindo esta resumida exposição das críticas deste importante autor, é necessária a exposição daquele que é para Dreyfus (1993) o mais intransponível obstáculo para a IA: a *criatividade*. Como imaginar ser possível a criação de um programa que, sendo em última análise uma complexa função lógico-matemática, seja capaz de adquirir regras novas? Como afirma Dreyfus, na verdade já foi dada uma prova matemática (e portanto *a priori*) das limitações inerentes a todos os sistemas formais, o Teorema de Gödel. Mas esta prova parece irrelevante para alguns pesquisadores da AI. De fato, o Teorema de Gödel implica logicamente que modificações no sistema no sentido de ampliar seus domínios necessariamente implicarão alterações nos axiomas, não dedutíveis do sistema anterior. Em outras palavras, não há programas logicamente estruturados que possam provocar nada parecido com uma mudança conceitual, uma introdução de teorema perfeitamente consistente com os outros do sistema mas não dedutível de seus axiomas. Não há função logarítmica (ou qualquer outra) que possa alterar a si mesma, portanto, não há programa que possa alterar a si mesmo. Aqui temos, como também assinalou Roger Penrose (1991), a mais fundamental distinção entre o processo mental humano e o

processamento computacional. Por conta desta cegueira teórica, Dreyfus (1993) compara os atuais pesquisadores da IA forte a alquimistas procurando transformar lata em ouro.

Mas aqui Dreyfus está sendo injusto com a Simulação Computadorizada e a IA. Como ele próprio reconhece, os alquimistas não conseguiram a pedra filosofal ou transmutar metais em ouro, mas nos legaram uma série de subprodutos e conhecimentos empíricos desorganizados. A IA tem nos proporcionado muito mais do que isso que Dreyfus admite. Ela tem nos proporcionado um corpo formalmente rigoroso e organizado de hipóteses e construído muitas vezes computadores desenhados somente para colocá-las em prática. Sem o esforço organizado de muitas das maiores mentes matemáticas e científicas de nosso tempo, não seria evidente para todos hoje que a mente humana tem atributos que vão muito além da possibilidade de formalização lógica. Somente hipóteses científicas levadas ao último nível de suas possibilidades são capazes de nos fornecer refutações espetaculares quanto as que têm se constatado no campo da IA forte, sobre *algumas* de suas pretensões. Só a ciência moderna pode nos revelar seu próprio limite, só ela pode nos conduzir com segurança às fronteiras da razão, e talvez, um dia nos revelar os limites da compreensão formal sobre a mente humana. Se é verdade que nossa mente não funciona como um computador boa parte do tempo, também é evidente que os tipos de inteligência simulados num computador são também apresentados pelo ser humano, o que indica que os processos subjacentes podem ser em grande medida semelhantes.

A questão da *qualia*

Outro filósofo que fez críticas que lembraram limites óbvios e intransponíveis para o computacionalismo foi Thomas Nagel. Em seu antológico artigo “*What is it Like to be a Bat?*”, Nagel (1980) nos lembra de forma bastante ilustrativa a verdade banal da irredutibilidade da experiência subjetiva, ou

experiência em primeira-pessoa (*first-person experience*). Em essência, seu argumento é que nenhuma quantidade de conhecimento descritivo em terceira-pessoa (ou seja, na forma de um investigador descrevendo a experiência de um outro sujeito), pode nos ajudar a experienciar como experimenta o mundo um morcego. Hipotetizamos que um morcego não tem visão, e que ele se orienta espacialmente emitindo sons e sentindo o tempo de resposta do obstáculo mais próximo. Ou seja, acreditamos que um morcego se orienta por uma espécie de sonar orgânico. Mas nós nunca seremos capazes de saber como é *ser* um morcego, como é *perceber* como um morcego e *decidir* como um morcego. Ele reconstrói um mapa cognitivo espacial da mesma natureza que o nosso? Ele se orienta por uma espécie de representação completamente distinta da espacial tridimensional? Ou com nenhum tipo de representação? Nenhum exercício de imaginação pode nos ajudar de fato. Nós nunca saberemos como outras pessoas sentem, e nunca teremos linguagem capaz de descrever adequadamente experiências subjetivas. Isso não significa que experiências conscientes não são reais, significativas, complexas, ricas e *altamente específicas*, únicas.

Esta questão é conhecida no debate filosófico como a questão da *qualia* (qualidade, singularidade), as qualidades fenomenológicas da consciência, como sentir dor, ver uma cor, sentir um sabor ou ouvir uma música. Esta qualidade da experiência subjetiva é irreduzível a explicações neurológicas ou processamento de informação. O melhor exemplo para ilustrar esta questão é o argumento desconcertantemente simples de Frank Jackson (1990). Se hipotetizarmos o surgimento de alguém em um futuro, onde o conhecimento neurofisiológico e físico estivesse completo, que comprovasse fazer discriminações de cores além do espectro visível, os neurocientistas seriam capazes de fazer uma descrição detalhada do funcionamento e estrutura de seu sistema nervoso, assim como do processamento da informação visual. No

entanto, como ninguém possui aquele cérebro, como ninguém é aquele homem, ninguém jamais saberá como é *ver*, *realmente experimentar*, estas cores que nunca ninguém viu. Portanto, a Física será sempre incompleta: a consciência é um domínio irreduzível.

Os argumentos de Thomas Nagel e Frank Jackson, são irrefutáveis. Eles nos lembram de uma fronteira intransponível para qualquer Psicologia filosófica ou científica futura: o domínio da *qualia*.

Rychlak e o pensamento dialético

Joseph Rychlak rejeita a tese de que a solução inatista do computacionalismo de Fodor representa um compromisso com uma visão kantiana da Psicologia. Para ele, as categorias kantianas não tem nada a ver com os receptáculos inertes de organização de idéias que ele identifica com a tradição lockeana. As categorias kantianas para Rychlak (1988) são “causas formais” da experiência, em virtude das quais a realidade percebida pelo sujeito é ativamente organizada. Assim, a nossa estrutura teórica formal irá condicionar o input que receberemos, ativamente, atribuindo sentido à experiência.

Rychlak (1988) faz críticas pertinentes à abordagem computacional da mente, que pretende ter resolvido o problema do comportamento humano orientado a metas, ou seja, pró-ativo, e com isso solucionado a questão teleológica em Psicologia. Ele se refere aqui inicialmente à famosa obra de Miller e colaboradores (1960), um dos marcos fundadores do Cognitivismo: “*Plans and the Structure of Behavior*”. Para estes autores, podemos definir um plano de maneira rigorosa como um processo hierárquico de seqüências de operações a serem executadas por um organismo, da mesma forma como um programa para um computador. Este nós conhecemos hoje como TOTE (test-operate-test-exit), um modelo cibernético de auto-regulação orientada a metas, ou feedback. A diferença aqui para Rychlak é que temos um modelo formal para “dar conta” do fenômeno

da intencionalidade do comportamento, não uma legítima aceitação da causa final. Temos causas formais e eficientes “dando conta” de uma formulação aceitável de parte dos aspectos pró-ativos do comportamento. Para os autores cognitivistas citados, intenção é uma parte incompleta de um plano cuja execução já tenha começado.

Rychlak questiona esta visão da atividade finalista humana, pois para ele esta deveria dar conta não da hierarquia de um plano de ação, mas da própria *definição* dessa hierarquia e desse plano. Caso remetêssemos a questão a planos e hierarquias maiores, estaríamos sempre somente transferindo o problema da legítima causalidade final para mais atrás, até termos que nos deparar com as metas e finalidades irreduzíveis (por exemplo, o plano de ir à faculdade, faz parte de uma meta mais elevada de terminar o doutorado, que faz parte de um plano mais extenso de formação profissional, que faz parte de uma meta mais básica de investigar profundamente certos problemas, que por sua vez precisa ser explicada sempre por uma hierarquia superior de metas).

Se um organismo está somente executando planos, então em qual sentido podemos falar de explicação teleológica? Só podemos falar de teleologia quando *formulamos* estes planos, *comparamos* planos diferentes e os *escolhemos*. A execução, assim como a execução de um programa, pode ser pensada em termos de feedback e causação eficiente, mas esta não é a questão para Rychlak (1988). Não temos aqui qualquer revolução em relação à imagem mecanicista de homem herdada do Behaviorismo, diz Rychlak. O comportamento continua a ser visto como explicado em termos de causa eficiente (impulsos neuronais) guiada pela causa formal do padrão do plano do “programa” (meta cognitiva). Mas onde está a verdadeira questão da pró-atividade, que é a *escolha* de planos e a *decisão de executar* o plano? Na imagem de homem do Cognitívismo como a vê Rychlak, em nenhum lugar. Já nos computadores é fácil: no *usuário*. É o programador que escolhe os planos e que

toma a decisão de rodá-los a maioria das vezes, e algumas poucas outras vezes nós, comuns usuários.

Em outro texto Rychlak (1986) ilustra bem esta questão da ausência de um verdadeiro comportamento teleológico em computadores. Ele sintetiza sua tese com a sentença: “*Computers do not predicate*” (Rychlak, 1986: 757). O comportamento orientado a metas, “intencional” do computador é de uma intencionalidade derivada, e derivada do *programador*. A diferença entre a suposta intencionalidade do computador e a verdadeira intencionalidade que é a do programador, como observa sarcasticamente Rychlak, não é problema para o sistema jurídico americano: as penalidades sobre malfuncionamento, danos e delitos cometidos por um software em execução recaem sempre sobre o programador, conforme legislação já em vigor nos Estados Unidos. Isto acontece porque o sistema legal presume que o verdadeiro agente do comportamento do sistema é a pessoa que o programou. Só a Psicologia, observa ironicamente Rychlak (1986), nega essa condição de predador universal ao ser humano.

Para não continuar nesta lastimável situação, Rychlak (1994) propõe para a Psicologia sua própria teoria. Em “*Logical Learning Theory*”, Rychlak apresenta a forma final de sua LLT e o resultado acumulado de décadas de pesquisa em seu suporte. Uma das alegações centrais da LLT é que o ser humano raciocina de duas formas básicas: a demonstrativa, sem questionamento das premissas assumidas, e a dialética, quando a indefinição entre premissas opostas e comparação entre elas. Para Rychlak (1994), é o pensamento dialético, que lida com as premissas que escolheremos para interpretar a realidade e as informações que receberemos, que é a raiz da liberdade subjetiva humana. É ao raciocinar dialeticamente que fazemos escolhas primevas sobre planos, metas e pressupostos. Lembrando o conceito kantiano de *dialética transcendental*, Rychlak advoga que idéias podem ser formuladas numa esfera transcendente, assim literalmente sendo capazes de rearranjar a realidade, ainda que

em detrimento do sujeito por causa das distorções resultantes. Uma vez que o *noumeno* é incognoscível, o que recebemos dos sentidos é informação. Este “input” que nos vem através das sensações não é somente ordenado pela formas cognitivas, mas pode potencialmente ter seu significado alterado por ele.

O problema do homúnculo e das três dimensões da mente

A última proposta deste trabalho é a de uma nova metáfora computacional para a Psicologia e a Ciência Cognitiva, não como um novo modelo para a mente, mas com o objetivo muito mais modesto de ajudar a ilustrar a divisão de fronteiras e tarefas desta Psicologia multifacetada e multicausal que surge no início do milênio. Para esta tarefa, devo convocar de novo uma personagem ridicularizada na Filosofia e na Psicologia: o *homúnculo*.

A metáfora do *homúnculo*, cuja origem remonta a Aristóteles, é mais um argumento reinterpretado de forma anedótica por Gilbert Ryle (1949). Trata-se de uma *reductio ad absurdum*. Supondo que alguém realmente representa um objeto internamente (como uma cadeira), como pode o cérebro lidar com esta representação? Certamente, esta representação está se dando em algum lugar do cérebro, e outro lugar do cérebro necessariamente tem que estar agindo sobre ela e a inspecionando. Esta outra parte é o homúnculo. Porque? Porque um problema idêntico agora se transfere para ele, o pequeno homenzinho dentro do cérebro. Se ele é a instância interna para a qual uma representação (de cadeira) deve se dar, para ele lidar com a informação trazida pela representação ele precisa representar a representação (de cadeira) de alguma forma. Mas esta representação a seu turno precisa de um homúnculo menor dentro do homúnculo para a qual ela está representada, e assim *ad finitum*. O paradoxo é interessante, o problema é a conclusão de Ryle: ela é a de que algo como *a representação mental não*

existe, pois leva a um regresso infinito na cadeia de causalidade.

É interessante ver como os pressupostos metafísicos assumidos por alguém podem levar as interpretações deste sobre as evidências para onde se quiser. Como uma pessoa pode chegar a pressupor que a representação mental não existe só porque não encontra a resposta de como ela é possível? Alguém realmente pode acreditar que não representa objetos na mente, ou que ao ler estas páginas não está representando mentalmente conceitos e os próprios signos lingüísticos? Como afirma Baars (1986), ninguém mais pode atualmente fazer tais afirmações. A Ciência Cognitiva é uma realidade, embora as representações mentais sempre tenham sido realidade desde o aparecimento do primeiro humanóide, e talvez também antes dele. Estes tipos de declarações ilustram o absurdo da atitude de tentar expulsar o fenômeno da consciência do mundo real, somente porque não há lugar para ele no mundo material mecanicista (que não existe mais nem na Física). Como nos lembra Searle (2000: 82):

“Qualquer tentativa de descrever a consciência, qualquer tentativa de mostrar como a consciência se encaixa no mundo em geral, sempre me parece inadequada. O que estamos deixando de lado é que a consciência não é apenas um aspecto importante da realidade. Em certo sentido ela é o aspecto mais importante da realidade, porque todas as outras coisas só tem valor, importância ou mérito em relação à consciência. Se valorizamos a vida, a justiça, a beleza, a sobrevivência, a reprodução, só as valorizamos como seres conscientes. Em discussões públicas, freqüentemente me pedem para dizer porque penso que a consciência é importante; qualquer resposta que se possa dar é sempre lamentavelmente inadequada, porque tudo que é importante é importante em relação à consciência.”

Creio que a interpretação adequada para o paradoxo do homúnculo é a mesma que a do paradoxo da regressão infinita da vontade (Ryle, 1949) e da intencionalidade dos sistemas computacionais. Vimos que a intencionalidade de um programa é derivada (Searle, 1992, Dreyfus, 1993). Um computador não atribui significado à informação. Seu *usuário* atribui. O paradoxo do homúnculo é uma prova de que não podemos reduzir a mente ao cérebro, não uma prova de que a mente não existe. Ou seja, o cérebro não representa algo para uma outra parte do cérebro, representa para a consciência. Para resolver este paradoxo, temos que postular das duas uma alternativa: ou a consciência é um fenômeno provocado pelo cérebro mas de ordem superior e portanto irreduzível a este (Searle, 1992; Sperry, 1993), ou a consciência é um fenômeno totalmente distinto do cérebro, que é a solução tanto de Descartes e Brentano (dualismo de substâncias) quanto do panpsiquismo (monismo de substância). De uma forma ou de outra, parece que não há alternativa para a Psicologia: temos que considerar o fenômeno central da vida psicológica e do universo conhecido: a *consciência*. E em virtude de suas categorias de intencionalidade primária, atribuição de significado, criatividade e escolha (*agency*), temos também que postular três níveis de análise irreduzíveis do fenômeno psicológico.

Isto já fazem até alguns materialistas como o filósofo Daniel Dennett (1978), que distingue três instâncias da explicação cognitiva: a *intencional*, que é o tipo de explicação e predição do comportamento que leva em conta desejos e crenças, mais exatamente metas e informação; a *design stance*, que poderíamos chamar de *formal*, que especifica os algoritmos (programas) que produzem o comportamento intencional; e por fim a *física*, que é o hardware, só relevante para explicar as superiores quando alguma coisa elétrica vai mal. É uma boa tripartição, mas insuficiente. O interessante é como o materialista Dennett acha que pode resolver o problema do homúnculo implícito na primeira instância: movendo a determinação da

intencionalidade da primeira instância para a segunda, você se livraria de um único homúnculo inteligente para obter um exército de idiotas (*army of idiots*), cuja única intencionalidade é ficar ativo ou inativo (0 e 1). Este é único caso da história da filosofia em que um filósofo acredita ter resolvido um problema multiplicando-o por dez bilhões. Isso nos mostra o quanto o materialismo também pode ser irracional.

David Marr (1982) ofereceu uma proposta de arcabouço muito semelhante para organizar os processos mentais. Ela consiste em três níveis de teorias. O nível mais alto é o *computacional* (que é preferível chamar de *funcional*, para evitar confusões), e contém a especificação do que precisa ser computado para que uma tarefa específica possa ser desempenhada. O nível intermediário é o *algoritmo*, que é o da linguagem de programação, é como os programas podem ser implementados, a representação para o input e o algoritmo de transformação. Por fim temos o nível do *hardware*, que é o da base física onde a representação e o algoritmo podem ser realizados de fato. O problema com o tipo de metáfora ou modelo de Marr, é que tal como o de Dennett, ele cai no paradoxo do homúnculo. Como observa Searle (1992), o que Marr está propondo trata o cérebro como se houvesse um agente dentro dele o usando para computar com ele: a tarefa da visão é descrita como um procedimento para transformar o input da imagem bidimensional que chega na retina numa descrição tridimensional do mundo externo como *output*. A dificuldade, diz Searle (1992), é: quem está vendo e atribuindo significado à descrição (ao *output*)? Assim, Searle acredita que todas as metáforas deste gênero invocam implícita e secretamente o homúnculo para tornar estas operações genuinamente computacionais.

Allen Newell (1982) é outro importante cientista cognitivo que invoca três níveis de explicação: o hardware, o programa e a intencionalidade. Este modelo é mais semelhante ao que este trabalho defende que deva ser proposto. E isto por um motivo simples: como observa Searle (1992), para

computadores que você compra numa loja não há o paradoxo do homúnculo, *porque o homúnculo é você*. Mas, ele afirma, se nós vamos supor que o cérebro é um computador digital, então seremos mais cedo ou mais tarde confrontados com a pergunta: “*And Who is the user?*” (p.214)

Não existe informação, a não ser para uma consciência. Tudo o que é codificado, pode ser recodificado e decodificado milhares de vezes, mas só ganhará significado *para uma consciência*. Desde a primeira vez em que tomei contato com as teses da IA forte, lembro-me de um filme da minha infância que era uma das poucas fitas de vídeo que tínhamos em casa quando surgiu este aparelho no início da década de oitenta. Seu nome era *Tron*, uma fábula infantil da Walt Disney onde os bits dentro do computador ganhavam vida e em momentos críticos perguntavam-se uns aos outros: “Você acredita no *usuário*?” Os homúnculos de Dennett (1978) me fazem lembrar destes bits e deste filme ruim.

Nós, seres humanos, não temos alternativa. Nós temos que acreditar no usuário porque *nós somos o usuário*. A consciência é o fenômeno psicológico primário. Não se podem ignorar os fenômenos mentais de primeira-pessoa e tratá-los como fenômenos de terceira-pessoa como a computação. Os sistemas só possuem uma intencionalidade derivada (Searle, 1992), eles dependem de um sistema intencional original, primário, que possa interpretá-los.

Por uma nova metáfora computacional

Assim proponho uma nova metáfora computacional, não como modelo para uma teoria da mente, mas como modelo de uma nova divisão de fronteiras da investigação psicológica. Imaginemos a seguinte situação metafórica. Um ser humano cresceu sozinho na floresta, e subitamente é colocado num quarto fechado, sem janelas, no qual seu único contato com o mundo é um computador plugado vinte e quatro horas por dia na internet. Ele só recebe informações do mundo externo de forma codificada pela rede, e só se comunica com o mundo lá fora da mesma

maneira. Ele precisa ter em seu computador um sistema operacional básico, capaz de permiti-lo fazer o primeiro acesso à Internet. A partir daí, ele precisa aprender a usar a máquina e o programa original. Pessoas há mais tempo na rede que ele, vão conseguir enviar mensagens visuais e auditivas para o seu computador o estimulando a mexer nele. Todo mundo, é visto através da tela do computador e ouvido através das caixas de som do computador. Progressivamente o incauto neo-informata vai aprender uma linguagem e os rudimentos de utilização de sua máquina. As pessoas com quem ele estabeleceu ligações de afeto vão sugerir que ele instale novos programas em sua máquina, para lhe permitir fazer mais coisas. Ele vai decidir quais instalar, mas no começo, eles serão instalados com base na confiança. Logo ele vai ver que tão logo instale um programa, há tarefas na rede que pode fazer muito mais rápido, sem a necessidade de sua intervenção permanente: uma vez instalado, é só colocar um programa para impedir invasões ou entradas em sites desagradáveis, ou para encontrar um caminho na rede enquanto ele conversa com uma amiga e assim por diante. Com o tempo, sua habilidade vai ficar tão ampliada, que ele vai passar a criar alguns programas, e alguns deles vai inclusive disponibilizar na rede, eventualmente, fazendo sucesso.

Enquanto for tendo experiências e aventuras virtuais, ele vai armazenar lembranças delas, fotos, imagens, textos, sons, músicas. Quando quiser lembrar do passado, ele vai resgatar estes traços. Ele não poderá armazenar tudo, mas tem mais coisa gravada em seu HD do que ele pensa: cookies, registros do sistema, traços de sua atividade na rede que permanecerão, fazendo que ele acesse mais rapidamente uma página por onde já passou, mesmo que conscientemente não se lembre mais dela. Mas nem tudo serão flores, como sabemos. Programas que entrem em conflito, podem fazer o computador entrar em pane total. Um super aquecimento na máquina pode prejudicar a execução dos programas. E aí ele vai ter dificuldade de ver, ouvir, rodar

processos automáticos. Se houver uma invasão de vírus, a máquina vai parar para removê-lo, ou pode mesmo ter parte de seu conteúdo apagado. Em todos estes casos, o hardware vai mandar uma comunicação para o sistema operacional, que vai invadir a tela com sinais visuais e as caixas com auditivos cada vez mais altos até que ele não consiga mais continuar suas atividades ordinárias, e pare tudo para resolver o problema. A maioria das vezes ele vai conseguir, e vai seguir em frente. Mas algumas vezes isso não será possível: o hardware poderá ficar irremediavelmente danificado e não funcionar mais corretamente. O sistema pode ficar tão corrompido que não conseguirá executar processos automáticos, ou os ficará executando sem parar sem que o usuário consiga fazê-lo parar. Por fim, quer tudo tenha saído bem quer não tenha, a memória começará a falhar, o monitor a tremer ou queimar, até que a vida útil do processador acabe e ele queime. Fim da história. O usuário não tem mais como se comunicar de nenhuma forma com aquele mundo virtual: não tem seus códigos, suas coordenadas, nem mais acesso aos traços mnemônicos.

Não é uma metáfora perfeita, mas é ilustrativa. Já sabemos que não podemos mais lidar com uma metáfora hardware-software, ela não funciona plenamente. Precisamos de uma metáfora *hardware-software-usuário*. Precisamos do homúnculo, mas precisamos dele fora da máquina. Da mesma substância que a máquina para poder interagir com ela, mas de ordem distinta, tão distinta do software quanto este é do hardware. Um quadro de Monet, quando vira um programa, não é um quadro de Monet. Ele só se torna um quadro de Monet novamente, quando decodificado de uma forma que nós possamos em seguida decodificar de forma significativa. Na tela, só existem pontos coloridos. No programa, uma série de instruções matemáticas ponto a ponto. Na nossa consciência, há um jardim, diferente e belo.

A consciência pode ser gerada pela atividade neuronal, mas também pode não ser. Isto pouco importa para a Psicologia científica, pois é metafísica. O que importa é

que *a consciência existe*. Ela precisa do cérebro para receber e decodificar as informações físicas, e para codificar fisicamente suas vontades transformando-as em ação corporal. Ela se serve do cérebro para executar ações automáticas, mas ainda assim precisa colocá-las em ação, como quando decidimos dirigir o carro até em casa, e só nos lembramos de novo da tarefa quando temos que abrir o portão. Os programas dependem da intencionalidade da consciência, porque senão as informações que eles manipulam não tem significado. Além disso, eles dependem da decisão de serem instalados ou apagados, embora algumas crenças e programas invasores sejam difíceis de apagar deixando rastros de auto-reinstalação. Programas novos, também são fruto da criação de um usuário, assim como nossas novas idéias e hipóteses. Nem todas os programas nós baixamos da rede, nem todas as nossas crenças são produtos unicamente da cultura. Nossa memória também é construtiva, não temos memórias perfeitas de tudo, só traços, a partir dos quais nós reconstruímos com a imaginação o acontecido. É claro que alguns eventos de importância, com a memória instantânea, podem ser gravados “completos” (som, imagem, dados) mas não teríamos memória suficiente para gravar desta forma tantas informações, por isso, assim como o computador, só gravamos algumas. O computador é como disse Howard Gardner (1996), uma metáfora adequada para explicar a execução de tarefas elementares e impenetráveis, como a percepção visual ou a análise sintática. Mas uma vez que rumamos para processos mais complexos e centrais “*como a classificação de domínios ontológicos e julgamentos referentes a cursos de ação rivais o modelo computacional se torna menos adequado*” (Gardner, 1986: 405). A metáfora hardware-software-usuário, também é útil para diferenciarmos dois tipos de motivação, a que surge na ausência de demandas dos softwares e do hardware, e a que surge para eliminá-las. A primeira é a do campo dos valores, a segunda, do campo respectivamente do desejo e da necessidade.

Estas demandas invadem nossa consciência e só desaparecem quando eliminadas ou dirimidas. Serve igualmente para ilustrar metaforicamente várias psicopatologias, como fobias, obsessão-compulsão, esquizofrenia, assim como lesões neurológicas e seus efeitos cognitivos.

A verdadeira utilidade desta metáfora no entanto, é ilustrar a nova divisão que se faz cada dia mais clara no campo da Psicologia. Em sua condição multicausal, a Psicologia encontra um campo explicativo na *Neuropsicologia*, que é a dimensão da relação hardware-software. Mas não podemos esperar explicações dedutivo-nomológicas (determinação de causas eficientes), somente explicações condicionais, danos ou upgrades no hardware não causam diretamente nenhum comportamento final, mas são condições de possibilidade dele: sem hardware, sem software, sem hardware, sem informação externa. Mas um hardware sem programa ou quem o instale e coloque para rodar não pode causar comportamento inteligente. O segundo campo explicativo é o da *Psicologia Cognitiva*, que é o campo do software e das suas relações com o usuário. Novamente aqui se coloca a condição de possibilidade, não de causa: ter um programa Word instalado não causa a aparição de um soneto de Shakespeare, mas é condição de sua possibilidade. Ter um Corel Draw instalado não causa uma figura da Mona Lisa, mas é condição de sua possibilidade. Por fim, o nível do usuário é o nível da consciência e das suas relações com os programas: é o campo explicativo da *Filosofia da Mente*, e guarda com a Psicologia Cognitiva um campo de intersecção.

Nesta nova configuração, a dimensão do usuário nos faz lembrar que qualquer metáfora computacional será sempre incompleta em relação ao ser humano, e sempre precisamos recorrer ao humano (mesmo que a um homúnculo) quando percebemos que não há nada no universo que possua ou possa representar algumas de nossas maravilhosas e inacreditáveis características, como a criatividade, a intencionalidade, a atribuição de sentido e

significado, a qualidade subjetiva, a vivência de valores e a liberdade.

Referências bibliográficas

- Baars, B.J. (1986). *The Cognitive Revolution in Psychology*. New York: Guilford.
- Castañon, G. (2006). John Searle e o cognitivismo. *Cien. Cogn.*, 8, 96-109. Disponível em www.cienciasecognicao.org.
- Dennett, D. (1978). Intentional Systems. In: Dennett, D., *Brainstorms*. Brighton: Harvester Press. pp. 03-23.
- Dreyfus, H. (1972). *What Computers Can't Do: The Limits of Artificial Intelligence*. New York: Harper & Row.
- Dreyfus, H. (1993). *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge, Massachusetts: The MIT Press.
- Fodor, J. A. (1975). *The Language of Thought*. Cambridge, Massachusetts: Harvard University Press.
- Gardner, H. (1996). *A Nova Ciência da Mente*. São Paulo: Editora da Universidade de São Paulo.
- Jackson, F. (1990). Epiphenomenal Qualia. Em: W. G. Lycan (ed.), *Mind and Cognition. A Reader*. Oxford: Blackwell. pp. 469-477.
- Marr, D. (1982). *Vision: A Computational investigation into the human representation of visual information*. San Francisco: W. H. Freeman.
- Miller, G., Galanter, E. & Pribram, K. (1960). *Plans and the Structure of Behavior*. New York: Holt, Rinehart & Winston.
- Nagel, T. (1980). What is it Like to be a Bat? Em: Block (Org.), *Readings in the Philosophy of Psychology*, vol. 1. London: Routledge & Kegan Paul.
- Neisser, U. (1967). *Cognitive Psychology*. New York: Appleton-Century-Crofts.
- Newell, A. (1982). The Knowledge Level. *Artificial Intel.*, 18, 87-127.
- Penrose, R. (1991). *A Mente Nova do Rei: computadores, mentes e as leis da Física*. Rio de Janeiro: Editora Campus.
- Putnam, H. (1961). Minds and Machines. Em: S. Hook (org.). *Dimensions of Mind*. New York: Collier. pp. 221-231.

Rychlak, J. (1986). Logical Learning Theory: A Teleological Alternative in the Field of Personality. *J. Personal.*, 54, 734-762.

Rychlak, J. (1988). *The Psychology of Rigorous Humanism*. New York: New York University Press.

Rychlak, J. (1994). *Logical Learning Theory: A Human Teleology and its Empirical Support*. Nebraska: University of Nebraska Press.

Ryle, G. (1949). *The Concept of Mind*. London: Hutchinson.

Searle, J. (1984). *Mente, Cérebro e Ciência*. Lisboa: Edições 70.

Searle, J. (1992). *The Rediscovery of the Mind*. Cambridge, Massachusetts: The MIT Press.

Searle, J. (2000). *Mente, Linguagem e Sociedade: Filosofia no Mundo Real*. Rio de Janeiro: Editora Rocco.

Sperry, R. (1993). The impact and promise of the cognitive revolution. *Amer. Psychol.*, 48, 878-885.