

John Searle e o cognitivismo

John Searle and the cognitivism

Gustavo Arja Castañón[✉]

Universidade Estácio de Sá (UNESA), Rio de Janeiro, Rio de Janeiro, Brasil

Resumo

A obra do norte-americano John Searle tem se tornado uma das maiores influências filosóficas na abordagem cognitivista da psicologia contemporânea, e começa hoje a rivalizar mesmo com a influência das obras de Jerry Fodor ou Noam Chomsky. Seu maior papel nas ciências cognitivas pode ser encontrado na ajuda ao resgate do conceito de consciência e ao ataque às teses da IA forte. Este ataque se divide em três frentes, uma para cada uma das três alegações centrais do computacionalismo. A primeira destas alegações seria 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”. Este artigo apresenta as críticas de Searle a estas teses, e mostra como suas posições acerca da mente, apesar de recentes, já influenciaram o cognitivismo, ajudando este movimento a retomar seu projeto original de se constituir como psicologia da consciência. © Ciências & Cognição 2006; Vol. 08: 96-109.

Palavras-chave: John Searle; cognitivismo; computacionalismo; consciência; inteligência artificial.

Abstract

The work of the North American John Searle is becoming one of the largest philosophical influences in the cognitivist approach of contemporary psychology, beginning to compete today even with the influence of the works of Jerry Fodor or Noam Chomsky. His largest role in cognitive sciences can be found in his help with the rescue of the concept of consciousness and in the attack to the theories of strong AI. This attack becomes separated in three fronts, one for each of the three central allegations of the computationalism. The first of these allegations would be that "the brain is a digital computer". The second, is that "the mind is a computational program". The third, is that "the operations of the brain can be simulated in a digital computer". This article presents Searle's critics over these theories, and also presents how their positions concerning the mind, in spite of recent, have already influenced the cognitivism, helping this movement to retake its original project of constituting itself as a psychology of consciousness. © Ciências & Cognição 2006; Vol. 08: 96-109.

Keywords: John Searle; cognitivism; computationalism; consciousness; artificial intelligence.

[✉] – **G.A.Castañón** é graduado em Psicologia (Universidade Estadual do Rio de Janeiro, UERJ) e em Filosofia (Universidade Federal do Rio de Janeiro, UFRJ), Mestre em Psicologia Social (UERJ) e doutor em Psicologia (UFRJ). Atualmente ministra cursos na graduação em Psicologia na UNESA, e cursa Mestrado em Lógica e Metafísica (UFRJ), tendo se dedicado nos últimos oito anos às investigações no campo de Filosofia da Psicologia. *E-mail* para correspondência: gustavocastanon@hotmail.com.

1. Introdução

O filósofo norte-americano John Searle é dono de uma das obras filosóficas mais importantes de nosso tempo. Nascido em 1932 e ligado à tradição analítica da filosofia contemporânea, começou sua produção filosófica abordando temas de filosofia da linguagem, que progressivamente se ampliaram para abranger tópicos de filosofia da mente e filosofia da ciência. Suas principais contribuições filosóficas foram o resgate da teoria fregeana dos nomes próprios, sua teoria da intencionalidade, sua teoria dos atos de fala e, mais recentemente, o aprofundamento de suas críticas ao computacionalismo, sua nova teoria da consciência e sua tese da construção da realidade social. Nesta revisão, serão apresentados resumos de sua nova teoria da consciência e de seu ataque ao computacionalismo, relacionando estas teorias com a abordagem cognitivista da psicologia contemporânea, evidenciando algumas influências atuais nesta e a compatibilidade das idéias de Searle com o espírito original da “revolução cognitiva”.

2. Computacionalismo, cognitivismo e ciências cognitivas

Vamos começar por uma definição de conceitos e distinção entre alguns termos que muitas vezes são usados como sinônimos, alguns deles centrais para este artigo. Dois destes são por exemplo os termos ‘cognitivismo’ e ‘psicologia cognitiva’. Como afirma Penna (1984), a psicologia cognitiva se pode conceituar tanto como um movimento doutrinário quanto como uma área específica de pesquisa. Baars (1986) observa que o termo ‘Psicologia Cognitiva’ é ambíguo, mas refere-se primariamente a uma “metateoria” que defende que através de observações empíricas podemos inferir constructos teóricos inobserváveis. Ele acredita que essa ambigüidade e confusão foram geradas porque a metateoria cognitiva surgiu no seio de uma disciplina psicológica também denominada *cognitiva* (o campo de estudo da memória, linguagem, percepção, pensamen-

to), mas que no entanto poderia ter surgido em qualquer outro campo da psicologia.

Aqui o termo ‘cognitivismo’ será usado para a referência ao movimento teórico da psicologia contemporânea, com seu modelo antropológico e metodológico (ou “metateoria”, Baars, 1986) enquanto o termo ‘psicologia cognitiva’ se refere à área de pesquisa específica. Enquanto área de pesquisa, a psicologia cognitiva pode se definir como *o estudo de como seres humanos percebem, processam, codificam, estocam, recuperam e utilizam informação*. É o estudo do processamento humano de informações. Enquanto movimento doutrinário na Psicologia, o cognitivismo foi definido por Penna (1984) como sendo marcado por cinco características principais. A primeira é a centralidade do conceito de *regra* para explicar o processamento cognitivo e o comportamento. A segunda o comprometimento com uma visão *construtivista* dos processos cognitivos. A terceira pela concepção do comportamento humano como orientado a *metas*. A quarta a imagem de um sujeito *ativo*, e não reativo como o da tradição positivista. Por fim, a quinta seria a recuperação do conceito de *consciência* na psicologia.

Outra distinção importante a ser efetuada é entre psicologia cognitiva e ciência cognitiva. Podemos definir ‘ciência cognitiva’ como um esforço multidisciplinar com fundamentação empírica para responder questões acerca da aquisição, armazenamento e utilização de conhecimento por parte do ser humano. A característica de esforço multidisciplinar é fundamental neste empreendimento, que consiste numa conjugação de esforços entre a psicologia cognitiva, a inteligência artificial, a neurociência, a lingüística e a filosofia da mente para explicar o processo de cognição humana. Apesar de a psicologia cognitiva ser a disciplina central de articulação destes esforços, ela não perde a especificidade de seu domínio nem a independência de seus métodos.

Por fim, devemos definir ‘computacionalismo’, contra o qual Searle se bate e do qual o cognitivismo deve ser distinguido.

Podemos chamar computacionalismo a crença de que os processos mentais consistem essencialmente em manipulação simbólica lógica, ou seja, computação. Também é conhecida pela abreviação CTM (*computational theory of mind*), e é associada classicamente a Jerry Fodor e seu clássico “*Language of Thought*”, de 1975. Na versão computacionalista de Fodor, os 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 confusão entre computacionalismo e cognitivismo tem várias origens. O principal delas é que, sem dúvida, a mais importante influência da revolução cognitiva na psicologia contemporânea foi o advento do computador. Isto se deu menos pelas Possibilidades futuras de simulação de processos cognitivos em máquinas do que -pela “metáfora computacional” (Neisser, 1967), a clareza conceitual que a distinção entre hardware e software permitiu à teorização sobre a mente e suas relações com o cérebro. Desde o surgimento da teoria matemática da “Máquina de Turing” até o desenvolvimento dos computadores de processamento paralelo, a chamada “metáfora computacional” tem sido a mais preciosa fonte de idéias para a psicologia cognitiva.

A primeira dessas idéias foi o sugestivo “Teste de Turing”, proposto pelo próprio pai da computação Alan Turing e que propunha que seria possível programar um computador de tal forma que seria impossível discriminar as respostas dadas por este a um interlocutor especialista (um psicólogo?) daquelas dadas por um ser humano. Tal teste logo se transformou na primeira expressão do credo da *tese forte* da inteligência artificial. Segundo este último, tal teste é plenamente apto a refutar qualquer um que duvide que um computador pode pensar: quando um especialista não for capaz de decidir se as respostas que estão aparecendo no monitor estão sendo dadas por um ser humano ou por um computador, afirma o credo, este especialista

estará diante do mesmo tipo de evidência da presença de pensamento do que quando está na presença de seu próprio filho, uma vez que consciência e inteligência não são entidades diretamente observáveis.

Assim, o materialismo filosófico ganhava finalmente uma versão um pouco mais sofisticada: livrava-se de um reducionismo fiscalista radical simplório sem aderir a um dualismo de substâncias. 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 seminal foi introduzida pelo filósofo Hilary Putnam (1961). Logo o materialismo pode se livrar dos estreitos limites behavioristas e partir para um tipo de investigação dos fenômenos mentais, criando o programa da *inteligência artificial*. Eles perceberam imediatamente que se conseguissem como psicólogos cognitivos descrever com precisão comportamentos e processos de pensamento de um organismo, poderiam ser capazes de projetar um computador e um programa que operasse de forma idêntica (simulação computadorizada). Mas o inverso também era verdadeiro. Caso os engenheiros de hardware e programação conseguissem projetar um computador e um programa que operasse de forma idêntica a um ser humano (inteligência artificial – IA), isto poderia nos levar a compreender o processo cognitivo humano. Era o nascimento da tese da IA forte.

Mas é necessário deixar claro desde o início que a maioria dos cientistas cognitivos de então – e essa tendência é hoje hegemônica – eram partidários da chamada *tese fraca* da Inteligência Artificial, e acreditavam que a construção de modelos computacionais que buscassem a simulação da inteligência humana serviria ao propósito mais modesto de revelar somente alguns aspectos mais automáticos da cognição humana, assim como evidenciar em que aspectos esta era *qualitativamente diferente* da “inteligência” computacional. E se isso já era verdade para a maioria dos pioneiros cientistas cognitivos, que dirá para o cognitivismo e a psicologia cognitiva. É preciso aqui ajudar a desfazer o mito de que para o cognitivismo o ser humano

é um computador neuronal e nada mais. Esta versão encontra força de propagação no fato de que, por decisão metodológica, a psicologia cognitiva simplifica artificialmente seus problemas colocando entre parênteses a influência nos processos cognitivos causada pelas emoções. Mas ainda assim, sua origem é a falsa associação do cognitivismo com a tese da IA forte, que sempre teve seu local de acolhida dentro das ciências cognitivas nas disciplinas das neurociências e da engenharia de computação, e nunca na psicologia cognitiva. De fato, não há qualquer psicólogo cognitivo ou cognitivista eminente partidário de tal tese. Como podemos facilmente constatar, nomes como Ulric Neisser (1967), George Miller (1985), Jerome Bruner (1997), Noam Chomsky (1981), Jerry Fodor (1998, 2001), Richard Mayer (1981), Howard Gardner (1996), Robert Sternberg (2000) e Aaron Beck (2000), se declaram sucessivas e repetidas vezes contra a IA forte e construíram teorias e carreiras que são a plena negação desta. No entanto, devemos lembrar que a maioria destes teóricos, caracteristicamente, reconhece a importância da metáfora computacional e da pesquisa em inteligência artificial como importante ferramenta da ciência cognitiva, aderindo à tese da IA fraca.

3. Breve panorama da obra de John Searle

A primeira obra de relevância filosófica de Searle foi seu artigo “*Proper Names*”, de 1958. Neste, Searle (1958/1996a) defende uma posição acerca da natureza dos nomes próprios condizente com as idéias de Gottlob Frege, e que consiste na tese segundo a qual todo nome próprio é associado a um determinado conjunto de descrições que determinam que objeto é designado por este nome. Em outras palavras, para Searle um nome é um designador indireto. Um nome próprio tem sentido não porque descreva características de um objeto (não descreve), mas porque está logicamente conectado com o conjunto das descrições definidas necessárias e suficientes para a descrição de um objeto particular.

O artigo citado rendeu ao jovem Searle, então com vinte e cinco anos, um precoce respeito da comunidade filosófica, particularmente daquela ligada à tradição analítica da filosofia ocidental. Mas foi com “*Meaning and Speech Acts*” (1962) e “*What is a Speech Act?*” (1965/1996b) que Searle começou a oferecer à filosofia contribuições verdadeiramente seminais. Sendo um desenvolvimento do trabalho de seu mestre John Austin, a teoria dos “*speech acts*” ou ‘atos de fala’ é a primeira de Searle a influenciar diretamente a psicologia cognitiva, particularmente a psicologia da linguagem (Sternberg, 2000). Em suma, Searle (1965/1996b) nos evidencia que um ato de elocução verbal é sempre também uma ação no mundo, ou seja, quando dizemos algo, sempre fazemos algo. Todas as declarações são atos de fala que necessariamente se enquadram em uma das cinco categorias seguintes: eles podem ser representativos (transmitem uma crença), diretivos (tentam levar alguém a fazer algo), comissivos (compromete o falante para uma ação futura, como por exemplo uma promessa), expressivos (expressam estados internos do falante) ou performativos (quando o ato de emissão da frase muda um estado de coisas intencional, como: “eu vos declaro marido e mulher”).

A teoria dos atos de fala implica num conceito mais fundamental, central na obra de Searle, que é o de intencionalidade. Em “*Intentionality: An Essay in the Philosophy of Mind*” (1983), John Searle nos oferece uma teoria geral da intencionalidade, isto é, daquela propriedade da consciência segundo a qual todo estado consciente se refere a algo de diverso da própria mente. Consciência é sempre consciência de algo, esse é o sentido de se dizer que a consciência é intencional, intenta algo. Quem ama, ama alguma coisa, quem representa, representa algo, e assim por diante. Em sua teoria da intencionalidade, Searle elabora conceitos hoje fundamentais na filosofia da mente, como *background* (ou “pano de fundo”), distinção geral dos tipos e conteúdos de atos intencionais, direção de ajuste e condições de satisfação.

Chega então à filosofia da mente, tendo trilhado o caminho natural que, partindo de um problema específico de filosofia da linguagem, passando pela teoria dos atos de fala e por uma teoria geral da intencionalidade, leva necessariamente a posições sobre a mente e a consciência. Estas começaram a ser expostas numa resposta antológica às pretensões da inteligência artificial, a metáfora do “quarto chinês”, o experimento mental mais célebre de nosso tempo. Antes no entanto de descrevermos esta e outras das idéias de Searle que nos interessam particularmente aqui por sua relação com o cognitivismo, cabe somente a citação das investigações à que Searle se dedicou nos últimos anos, combatendo alguns modismos anti-realistas de nosso tempo em “*The Construction of Social Reality*” (1995). Nesta obra, Searle se dedica à elaboração de uma filosofia das ciências sociais e a uma vigorosa defesa do realismo, contrapondo a idéia óbvia de que construímos de fato a realidade social, pois esta é uma criação humana, à idéia absurda de que construiríamos socialmente a realidade, mesmo em seus aspectos físicos.

4. Searle e a redescoberta da consciência

Em “*The Rediscovery of the Mind*”, de 1992, Searle apresenta sua rejeição tanto do materialismo quanto do dualismo. Ao dualismo, rejeita por considerar que a consciência é um fenômeno natural, ou seja, ela é de alguma forma gerada a partir da atividade cerebral. Ao materialismo, rejeita por considerar evidente que apesar disto a consciência não é redutível a um fenômeno físico. Em suma, o problema do materialismo é a negação da consciência: embora os materialistas não tenham coragem de assumir diretamente que estão afirmando que a consciência não existe, eles a redefinem de forma a que ela não mais se refira a estados mentais internos, qualitativos e subjetivos, mas sim a algum fenômeno de terceira pessoa.

Searle (1992) demonstra que respostas tais como que a consciência é comportamento do corpo, ou estados computáveis do cérebro, ou processamento de informações ou estados

funcionais de um sistema físico, simplesmente não funcionam. A consciência é um fenômeno interno, qualitativo, subjetivo e de primeira-pessoa: qualquer explicação da consciência que deixe de lado estas características, pode ser explicação de qualquer coisa, menos da consciência. O argumento que nos dá Searle (1992: 121-122) para ilustrar porque a consciência é um fenômeno que não se submete a uma explicação de caráter redutor, ou seja, uma explicação que reduz o fenômeno explicado como aparência de um fenômeno real que o constitui, é persuasiva:

“Parte da questão da redução no caso do calor era distinguir entre o aparecimento subjetivo deste por um lado e a realidade física subjacente no outro. Realmente, é uma característica geral de tais reduções o fenômeno ser definido em termos da 'realidade' e não em termos do 'aparecimento'. Mas nós não podemos fazer este tipo de distinção de aparecimento-realidade para consciência porque a consciência consiste nestes próprios aparecimentos. Quando o aparecimento é a questão não podemos fazer a distinção aparecimento-realidade porque o aparecimento é a realidade.”

Em outras palavras, o padrão das reduções eliminatórias é mostrar que o fenômeno reduzido é apenas uma ilusão. No entanto, no caso da consciência, esta “ilusão” é a realidade a ser explicada. Como afirma Searle (1992), quando me parece que estou consciente, então estou. Se estou iludido que estou consciente, então estou consciente. Quando um materialista diz que a consciência é uma ilusão que o cérebro se esforça arduamente em manter, ele parece se esquecer que essa “ilusão”, a partir da qual conhecemos tudo o que existe, é o fenômeno a ser explicado, e não processos materiais subjacentes. Como coloca Searle (2000: 59):

“... a consciência é diferente dos pontos porque eu posso ter a ilusão que o sol está se pondo por trás das

montanhas quando na verdade ele não está fazendo isso. Mas não posso, da mesma maneira, ter a ilusão da consciência sem estar consciente. A 'ilusão' da consciência é idêntica à consciência."

Não resta dúvida que este argumento é uma variação do cogito cartesiano. Searle (1992) lembra que a consciência é um objeto que apresenta problemas únicos à investigação, pois não podemos observá-la como observamos outros objetos, nem podemos fazer a distinção entre a observação e o objeto observado. No entanto, a consciência também apresenta facilidades únicas para a investigação, porque nós não experimentamos tão diretamente nenhum outro objeto no universo. Filosoficamente, diz Searle, podemos estabelecer doze de suas características.

A primeira é a de que ela se manifesta em um número estrito de *modalidades*. Estas envolvem o ver, o cheirar, o ouvir, o tocar, o sentir o gosto, o senso de equilíbrio, o sentir a temperatura, a propriocepção, o pensar em palavras, imagens e outras formas não verbais ou espaciais, o sentir prazer e desprazer físico e o sentir prazer e desprazer sem conexão com sensações corporais.

A segunda é a unidade, a consciência chega a nós como *fenômeno unitário*, unificado, não importa a quantidade de estímulos e pensamentos que estejamos tendo neste momento. A terceira é a *intencionalidade* no sentido de Brentano e Husserl, a consciência nos dá acesso a um mundo diferente dela mesma. A consciência é sempre consciência de algo, de duas formas estruturalmente distintas: ou representamos as coisas como acreditamos que elas são, e essas são nossas crenças, ou as representamos como gostaríamos ou temeríamos que elas fossem, e esses são nossos desejos (que usualmente chamamos de intenções, mas obviamente não são nossos únicos estados intencionais) e medos. Searle (1992) chama o primeiro modo intencional o modo cognitivo, e o segundo o modo volitivo.

A quarta característica da consciência para Searle (1992) é a *subjetividade* ontológica, a *qualia* irreduzível, a intransferibilidade de um estado consciente. 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; Nagel, 1980), as qualidades fenomenológicas da consciência, como estar sentindo dor, vendo uma cor, sentindo um sabor ou ouvindo uma música. Esta qualidade da experiência subjetiva é irreduzível a explicações neurológicas ou processamento de informação.

A quinta é a da *interdependência entre consciência e intencionalidade*, tese postulada por Searle de que só um ser que pode ter estados intencionais conscientes pode ter estados intencionais de fato, e que cada estado intencional inconsciente é ao menos potencialmente consciente (1992:132). A sexta característica é que estados conscientes são sempre *estruturados*, no sentido gestáltico da palavra. Estruturamos nossas experiências conscientes de forma instantânea em todos coerentes contextualmente. Um sétimo aspecto das experiências conscientes é que elas vêm sempre com vários graus de *familiaridade*, do muito familiar de nosso quarto ao muito alheio de um quarto onde você nunca entrou.

A oitava característica da consciência é o que Searle (1992) denomina *overflow*. O "transbordamento" a característica que tem os estados conscientes de cada experiência só ganhar sentido com experiências que estão além delas mas a ela conectadas por contigüidade. Um pensamento sempre liga a outros circundantes e assim indefinidamente. Uma nona característica é a que Searle caracteriza como semelhante à *distinção entre o centro e a periferia*. Aqui a metáfora da lanterna se torna necessária. A consciência é sempre um processo focado em alguns conteúdos enquanto uma grande gama de experiências

também conscientes estão sendo mantidas fora do centro do foco, como uma lanterna a iluminar parte do quarto enquanto o resto está parcialmente visível mas fora da atenção. Os exemplos clássicos são a sensação da sola de seu pé agora, ou talvez uma ligeira sede que, apesar de estar fazendo parte de sua consciência global, não ocupavam até este momento o foco de sua atenção.

A décima característica da consciência para Searle (1992) são suas *condições fronteiriças*, que ele descreve como um sentido de nossa própria situação relativa a tempo e espaço. A décima-primeira é o *mood*, que poderíamos traduzir como *humor*. Searle acredita que de maneira global podemos nos sentir felizes ou deprimidos, sem que esta condição esteja intencionalmente ligada a nenhum objeto (nenhum objeto ou situação específica nos deixa triste ou alegre). Anos depois Searle (2000) reviu sua posição excluindo o humor da lista de características estruturais da consciência. De fato, podemos pensar o humor como sendo ligado diretamente a crenças ou desejos inconscientes. Por fim, chegamos à última característica elencada por Searle, a dimensão prazer-desprazer de todos os estados conscientes, que são sempre prazerosos ou desprazerosos em algum nível.

O que fica depois da descrição de Searle da estrutura da consciência, é um sentido de certa impotência. Estamos diante do mais essencial e ainda assim complexo dos fenômenos. A consciência tem em si as características mais únicas de todo o universo conhecido, e ainda assim, as mais familiares. O que pode a ciência de terceira-pessoa, a ciência dos objetos nos dizer sobre a consciência? A resposta de Searle é simples, e é a mesma de Edmund Husserl: *nada*.

5. Searle e a crítica ao computacionalismo: é a mente um programa?

Seguindo a exposição iniciada no subitem acima, é preciso chamar a atenção para o fato que se pensar é manipular símbolos através de regras puramente formais e sintáticas, então a semântica, o significado das representações, não tem lugar na expli-

cação psicológica (Bem e de Jong, 1997). Sobre o que quer que sejam que as representações mentais, isso não determina sua manipulação e portanto o funcionamento de um sistema formal. O tipo de operação computacional que gera uma imagem da *Mona Lisa* num monitor, ou que faz ficarem pretas as partes do visor de cristal de uma calculadora onde vemos um número *quatro*, não tem nada a ver com o significado que estas imagens representam para um usuário. Searle (1984) foi o filósofo que ilustrou de maneira mais evidente esta questão, com sua já bastante conhecida metáfora do quarto chinês.

Em essência sua metáfora é a seguinte. Suponha que um matemático programador americano que não sabe absolutamente nada de chinês mandarim 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 uma frase em chinês pela manhã para passar o dia aplicando as regras nela, e conseguir enfim chegar à indicação no fim do dia dos ideogramas que ele deveria entregar como resposta. 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, produzindo ao final da aplicação outros símbolos como output. Assim como no teste de Turing, os ideogramas que ele recebe são perguntas sobre uma história, e os ideogramas que ele entrega são respostas adequadas a estas perguntas, o que leva um avaliador chinês que analisa as respostas a afirmar que ele compreende perfeitamente o texto (que ele nem leu). A pergunta é simples, e também a resposta. Ele teria compreendido o significado das frases? Obviamente não. Ele teria processado informação de acordo com regras lógicas? Obviamente sim. Logo, uma coisa independe da outra; logo, 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.

É importante aqui, neste momento, aproveitar para esclarecer uma das maiores confusões em relação à *psicologia cognitiva* e ao *cognitivismo*. É fato que em “*Minds, Brains and Science*” e “*The Rediscovery of the Mind*”, Searle critica o cognitivismo e as ciências cognitivas. Mas trata-se de um equívoco. Os argumentos de Searle não atingem o cognitivismo, nem sequer as teses atuais de Fodor, eles atingem sim a tese da *inteligência artificial forte*, defendida hoje por escassos setores das ciências cognitivas (graças estes mesmos argumentos e às críticas de autores como Nagel, 1980 e Dreyfus, 1993, entre outros). O argumento do quarto chinês é um argumento contra o teste de Turing. Fodor não reduz o pensamento a operações lógicas, justamente porque exclui dele o problema semântico, para o qual ainda procura solução (Fodor, 1998).

Assim, 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. O que se precisa determinar, é que regras ele usa e que símbolos ele armazena. O nosso objeto de estudo em psicologia cognitiva portanto, não é o significado da experiência como queriam os humanistas, mas as regras do pensamento (como a gramática transformacional de Chomsky) e as representações mentais de todos os tipos (como linguagem ou imagens).

Fodor (1991) transforma estas questões num princípio conhecido como *solipsismo metodológico*. O aspecto da mente que pode ser estudado é o puramente sintático. Portanto, a referência das representações ao mundo exterior está além dos poderes explicativos da psicologia cognitiva. Se uma paciente *acredita* em duendes que vivem em Mauá, isto *causa* nela o desejo de

vê-los, o que a *leva* a viajar à Mauá e *procurá-los* pelas matas. Quer existam ou não duendes, suas representações deles causaram e explicam seu comportamento. Não é necessário portanto nada externo ao sujeito psicológico para explicar o comportamento: somente as informações que ele recebe, as representações que tem e as regras que aplica para manipulá-las. É importante enfatizar que este é um princípio metodológico somente, não ontológico, não se nega a existência nem a importância do ambiente, se nega somente que ele seja termo de explicações psicológicas.

É impossível para o cognitivista fazer semântica, diz Fodor (1991), é impossível o acesso objetivo aos significados das representações (que permanecem no âmbito do sujeito), mas dizer isto é obviamente muito diferente do que afirmar o absurdo de que representações mentais não têm propriedades semânticas. O problema é que estas não são acessíveis à investigação científica. Aqui, mais do que em qualquer lugar da extensa obra de Fodor, se torna evidente que ele não é um computacionalista radical, e que as críticas de Searle relativas ao problema do significado e da *qualia*, não se aplicam a Fodor.

Fodor (1998, 2001) está perfeitamente consciente dos limites do Funcionalismo e de sua tese da linguagem do pensamento. A principal limitação para ele é a que se refere ao problema que Searle (1984) aponta, o problema do significado das representações. A concepção estrita-mente funcionalista do significado é holista, focada exclusivamente nas relações causais entre um símbolo e outro, permanecendo puramente interna ao sistema simbólico. É claro que esta é uma posição insustentável. Se uma palavra define seu significado pelas relações que mantém com outras, isto só pode funcionar se algumas delas tiverem significado estabelecido. As representações metaforicamente se igualariam assim a espelhos numa casa de espelhos sem visitantes, onde um espelho refletiria outro que refletiria outro que refletiria o primeiro *ad finitum*.

6. Searle e a crítica ao computacionalismo: é o cérebro um computador digital?

Assim, partindo da realidade, irredutibilidade e centralidade da consciência para explicar os fenômenos mentais, Searle (1992) aprofunda sua crítica ao computacionalismo com dois argumentos tão simples e devastadores quanto o mais antigo e conhecido argumento do quarto chinês. No entanto, mais uma vez aqui vemos que Searle não leva em consideração em sua obra a distinção entre o cognitivismo como movimento psicológico, as ciências cognitivas como empreendimento multidisciplinar e a inteligência artificial, tomando todos como expressões do computacionalismo. Feitas estas considerações, arrisco-me a afirmar aqui que sua obra “*The Rediscovery of the Mind*” marca o fim do projeto da inteligência artificial e da teoria computacional da mente clássica.

O ataque de Searle ao computacionalismo se divide em três frentes, uma para cada uma das três alegações centrais desta abordagem da mente. A primeira destas alegações seria para Searle (1992) 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”. Já vimos no item anterior o argumento de Searle contra a segunda tese central do computacionalismo, argumento que evidencia que a dimensão sintática não é suficiente para explicar o que faz a mente. Veremos agora como ele ataca a primeira e a terceira premissas desta abordagem da mente.

A terceira das teses centrais do computacionalismo é segundo Searle (1992) de que “as operações do cérebro podem ser simuladas em um computador digital”. Esta é eliminada por ele de forma desconcertantemente simples. Sua resposta é: *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.

Assim sobra a primeira alegação, a central, a fundamental tese do computacionalismo, de que “o cérebro é um computador digital”, a única que poderia trazer ainda algum alento a IA. Mas Searle (1992) aqui oferece a mais desconcertante de suas teses. Analogamente à questão 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*. Em suas palavras (Searle, 1992: 208):

“Seguindo a definição de computação padrão oferecida em livros-texto, é difícil ver como evitar os seguintes resultados:

- 1. Para qualquer objeto, existe alguma descrição de tal objeto tal que sob esta descrição o objeto é um computador digital.*
- 2. Para qualquer programa e para qualquer objeto suficientemente complexo, existe uma descrição deste objeto sob a qual este está implementando tal programa. Então a parede atrás de minhas costas, por exemplo, está neste exato momento implementando o programa Word, porque existe nela algum padrão de movimentos moleculares que é isomórfico com a estrutura formal do Word. Mas se a parede está rodando o Word, então esta é uma parede grande o bastante para estar implementando qualquer programa incluindo qualquer programa implementado no cérebro.”*

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”:

“A múltipla realizabilidade de processos computacionalmente equivalentes em meios físicos diferentes não é somente um sinal de que os processos são abstratos, mas também de que eles não são intrínsecos ao sistema. Eles dependem de uma interpretação de fora.” (Searle, 1992: 209)

Assim, com esta inversão explicativa e complementação implacável do argumento do quarto chinês, Searle demonstra que, não só as crenças básicas do computacionalismo são falsas, como possivelmente carecem de um sentido preciso. Qualquer coisa pode ser interpretada como um estado computacional, você pode considerar as moléculas da parede de sua casa como um programa implementado (depois da trilogia “Matrix”, até se fundaram seitas adolescentes que acreditam que estamos vivendo numa realidade virtual completa). *Mas a consciência não.* 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.

Searle (1992) acusa o computacionismo de ser uma versão disfarçada da falácia do homúnculo, que seria endêmica em todo o campo do cognitivismo (e aqui isto inclui a psicologia cognitiva). Ele sempre trataria 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 como o problema do homúnculo, porque *o homúnculo aqui somos nós.* A “intencionalidade” do computador é a nossa, porque ela é somente derivada. 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. Então o computador, através de uma série de estágios elétricos, transforma novamente os sinais elétricos em signos (na tela, numa folha impressa) nos quais a informação está codificada, para que um agente externo (que pode ser o próprio original – quando você registra informações para se lembrar posteriormente – ou outra 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).

7. Searle e o Cognitivismo

Como vimos, quando Searle (1984, 1992) critica o “cognitivismo”, ele na verdade está criticando o computacionalismo. Searle parece não reconhecer a existência de uma psicologia cognitiva autônoma, independente da inteligência artificial e das neurociências. Aqui temos um padrão comum com as críticas ao cognitivismo vindas da tradição humanista (Rychlak, 1988; Dreyfus, 1993), identificar cognitivismo com ciência cognitiva, e esta última com inteligência artificial servida por experimentos paralelos em neuropsicologia para colocar em teste com seres humanos os modelos construídos pela AI.

Esta segunda parte da crítica é totalmente falsa, e cada vez mais nos últimos anos, em que temos assistido a emergência da psicologia cognitiva como pólo de integração e carro chefe das ciências cognitivas (Eysenck e Keane, 1994; Sternberg, 1992), fonte principal de idéias e teorias para a simulação computadorizada e a neuropsicologia. Searle parece não conferir a menor importância ou significado à psicologia cognitiva, pois caso

conferisse, provavelmente saberia que suas posições sobre a realidade, irredutibilidade e centralidade da consciência para a explicação dos fenômenos mentais é compartilhada pela maioria dos principais nomes do cognitivismo contemporâneo, para os quais sua obra se tornou uma importante referência (Beck, 2000; Bruner, 1997; Gardner, 1996; Sternberg, 2000) e tem sido incorporada progressivamente pelo cognitivismo.

Como afirma George Miller (1985), um dos ícones do movimento cognitivista, o resgate do conceito de *consciência* é uma das características centrais da imagem de homem promovida por esta abordagem. A seguinte passagem ilustra bem esta postura:

“Psicólogos que adotam a consciência como o problema constitutivo de seu campo precisam rejeitar pouco do que se passa por psicologia hoje. Eles podem aceitar o inconsciente, por isto definir as fronteiras da consciência. Eles podem aceitar a análise do comportamento, por isto prover a evidência para processos conscientes. Eles podem aceitar estudos de crianças e animais, por eles revelarem o desenvolvimento da consciência. Eles podem aceitar simulação computadorizada, por isto ilustrar a lógica dos processos conscientes. Eles podem aceitar a atribuição social, por isto dar forma a nossa consciência dos outros. A fé central é que a consciência é um fenômeno natural e que a disciplina – eventualmente a ciência – responsável por compreender isto pode ser chamada psicologia.” (Miller, 1985: 42)

Assim o que vemos é que o cognitivismo se caracteriza como uma psicologia da consciência, como também defende Jerome Bruner. No final de sua carreira, Bruner (1997) passou a dar ênfase a aspectos que ele considerava esquecidos pela revolução cognitiva, defendendo que originalmente era o *significado*, e não o processamento de informações, o objeto central deste movimento. Defendendo o conceito de

pró-atividade e intencionalidade da consciência contra o computacionalismo, Bruner se alinha explicitamente a Searle (1997: 21) contra o computacionalismo característico das ciências cognitivas nos anos 60 e 70. O ser humano não é para o cogniti-vismo uma bola de bilhar reativa num universo mesa-de-bilhar newtoniano-meca-nicista. Ele é um foco de atividade do universo. Busca ativamente metas, constrói ativamente suas estruturas cognitivas, atribui ativamente significado.

Assim também pensa Aaron Beck (2000), maior nome da terapia cognitiva, outro que se alinha inclusive com as críticas mais recentes de Searle ao computacionalismo (Beck, 2000: 47) assim como à sua nova teoria da consciência (Beck, 2000: 55). Para ele,

“... a postura filosófica da teoria cognitiva sobre a questão do ‘livre-arbítrio’ reconhece a cognição como um mecanismo que pode, em parte, ser determinado ou controlado por variáveis externas. Contudo, ao mesmo tempo, a natureza da consciência humana inclui o potencial para causalidade e criatividade. Na verdade, sem esse potencial, não haveria novas teorias científicas das quais derivar hipóteses testáveis para pesquisa científica empírica.” (Beck, 2000: 46)

Assim como Searle, o cognitivismo não entende a consciência como um epifenômeno descartável, que uma vez excluído em nada alteraria as seqüências de comportamento a serem efetuadas por um organismo, conforme concebiam filósofos behavioristas e positivistas que ousavam abordar esta questão. No cognitivismo, a consciência não é o fantasma na máquina de Ryle (1949). Nomes como Neisser (1967) e Sperry (1993) consideram a consciência uma propriedade emergente da atividade cerebral que, no entanto, possui propriedades que não se reduzem às propriedades desta atividade (Sperry, 1993). De resto, cabe aqui lembrar um argumento oferecido por Penna (1984) em favor da relevância da consciência como

fenômeno biológico, contra os pouco verossímeis argumentos de Ryle e dos behavioristas. Neste, lembra ele que a tese behaviorista da irrelevância da consciência não se compatibiliza com o fato de sua preservação ao longo do processo evolutivo, já que segundo a teoria da evolução a vida tende a descartar as formações emergentes inúteis e desnecessárias. Se a consciência fosse um simples epifenômeno, seu destino já teria sido o desaparecimento.

Roger Sperry (1993), neuropsicólogo vencedor do prêmio Nobel com seu famoso estudo sobre o funcionamento cognitivo de pacientes com hemisférios cerebrais separados, é um dos maiores nomes do cognitivismo. Em sua obra procurou formular uma nova posição sobre a relação mente-corpo – que como vemos aqui, é bastante semelhante ao naturalismo não-materialista e não-reducionista de Searle – e levar o interacionismo característico do cognitivismo um passo a frente. Sperry não foge das dificuldades em afirmar uma tese sobre o que a mente é, nem sobre porque poderíamos falar de dualismo sem falar de dualismo ontológico (ou de substâncias). Partindo do pressuposto holista de que “o todo é mais que a soma de suas partes”, ou seja, de que os *todos* apresentam propriedades irreduzíveis às propriedades das partes que o constituem, Sperry, de forma semelhante a Searle (1998) apresenta a consciência como uma propriedade emergente da atividade cerebral, que portanto adquire propriedades distintas daquela. Neste sistema interacionista, o caminho da causação entre o todo (a mente) e as suas partes constituintes (os neurônios) é bi-direcional. No entanto ele é claro em afirmar que acredita que este “novo mentalismo” não é um novo tipo de dualismo ontológico:

“...em vista de equívocos comuns, ... vale repetir que o tipo de mentalismo defendido aqui não é dualista no sentido filosófico clássico de dois domínios diferentes e independentes de existência. Em nossa nova síntese ‘macromental’ ou ‘holomental’, estados

mentais como propriedades emergentes de estados mentais causam comportamento mas não são dualistas, porque são inextricavelmente ligados aos processos cerebrais que os geraram. Desta forma, estados mentais não podem existir separados de um cérebro ativo. Ao mesmo tempo, estados mentais não são o mesmo que estados cerebrais. Os dois diferem da mesma forma que uma propriedade emergente dinâmica difere de sua infra-estrutura componente. É característico de propriedades emergentes o serem notavelmente inusitadas e mesmo fantástica e inexplicavelmente diferentes dos componentes dos quais elas são feitas.” (Sperry, 1993: 06)

É muito difícil concluir até que ponto teorias interacionistas como as de Sperry (1993) ou Searle (1998) representam uma posição clara sobre o problema mente-corpo, como as doutrinas materialista e behaviorista monista certamente representam (embora sejam inverossímeis). Não é claro como estes autores podem acreditar ou justificar que de uma entidade material, o cérebro, possa se gerar como “propriedade emergente” algo que é absolutamente distinto deste, com propriedades “fantásticas e inexplicáveis”, únicas em todo o universo conhecido. Aqui, tudo o que podemos concluir é que dois mil e quinhentos anos de filosofia, cento e vinte cinco de psicologia, e cinquenta de ciência cognitiva, não fizeram muito pelo esclarecimento deste que é um dos maiores problemas da Filosofia e mistérios para o ser humano.

8. Conclusão

Este artigo apresentou argumentos de Searle que parecem indicar que o computacionalismo tradicional e a IA forte estão condenados ao desaparecimento. Ao mesmo tempo, procurou esclarecer um equívoco comum que identifica computacionalismo a cognitivismo, e que é herdado pelo próprio Searle. Ao contrário de um

enfraquecimento do cognitivismo, as críticas de Searle (1992) ao computacionalismo tem como resultado um reforço ao realinhamento do Cognitivismo ao seu projeto original de se constituir como autêntica psicologia da consciência, e oferece a este movimento um pequeno arsenal de posições filosóficas que pode o tornar muito mais coerente ontologicamente. A centralidade da consciência para a compreensão do comportamento e dos processos cognitivos é marca distintiva tanto do pensamento de Searle quanto do cognitivismo como abordagem da psicologia. Julgo que foram apresentados aqui motivos suficientes para crer que o tempo se encarregará de transformar a incipiente influência da filosofia de Searle neste movimento, em um casamento indissociável.

9. Referências bibliográficas

- Baars, B.J. (1986). *The Cognitive Revolution in Psychology*. New York: Guilford.
- Beck, A. e Alford, B. (2000). *O Poder Integrador da Terapia Cognitiva*. Porto Alegre: Artmed.
- Bem, S. e De Jong, H. (1997). *Theoretical Issues in Psychology: An Introduction*. Londres: Sage Publications.
- Bruner, J. (1997). *Atos de Significação*. Porto Alegre: Artes Médicas.
- Chomsky, N. (1981). *Regras e Representações*. Rio de Janeiro: Zahar Editores.
- Dreyfus, H. (1993). *What Computers Still Can't Do: A Critique of Artificial Reason*. The MIT Press: Cambridge, Massachusetts.
- Eysenck, M. e Keane, M. (1994). *Psicologia Cognitiva: Um Manual Introductório*. Porto Alegre: Artmed.
- Fodor, J.A. (1975). *The Language of Thought*. Cambridge: Harvard University Press.
- Fodor, J.A. (1991). Methodological solipsism considered as a research strategy in Cognitive Psychology. Em: Boyd, R. e Gasper, P. (Orgs.). *The Philosophy of Science*. p. 651-669. Cambridge: MIT Press.
- Fodor, J.A. (1998). *Concepts: Where Cognitive Science Went Wrong*. Oxford: Oxford University Press.
- Fodor, J.A. (2001). *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*. Cambridge: MIT Press.
- Gardner, H. (1996). *A Nova Ciência da Mente*. São Paulo: Editora da Universidade de São Paulo.
- Mayer, R. (1981). *The Promise of Cognitive Psychology*. São Francisco: W. H. Freeman & Company.
- Miller, G. (1985). The Constitutive Problem of Psychology. Em: Koch, S. e Leary, D. (orgs.). *A Century of Psychology as Science*. (pp. 40-46). New York: McGraw-Hill Book Company.
- Nagel, T. (1980). What is it Like to be a Bat? Em: Block (Org.), *Readings in the Philosophy of Psychology*, vol. 1. Londres: Routledge & Kegan Paul.
- Neisser, U. (1967). *Cognitive Psychology*. New York: Appleton-Century-Crofts.
- Penna, A.G. (1984). *Introdução à Psicologia Cognitiva*. São Paulo: Editora Pedagógica e Universitária - EPU.
- Putnam, H. (1961). Minds and Machines. Em: Hook, S. (org.). *Dimensions of Mind*. (pp. 221-231). New York: Collier.
- Rychlak, J. (1988). *The Psychology of Rigorous Humanism*. New York: New York University Press.
- Ryle, G. (1949). *The Concept of Mind*. Londres: Hutchinson.
- Searle, J. (1962). Meaning and Speech Acts. *Philosophical Rev.*, 71.
- Searle, J. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.
- Searle, J. (1984). *Mente, Cérebro e Ciência*. Lisboa: Edições 70.
- Searle, J. (1992). *The Rediscovery of the Mind*. Cambridge: MIT Press.
- Searle, J. (1995). *The Construction of Social Reality*. 1ª ed. New York: Free Press.
- Searle, J. (1996a). Proper Names. Em: Martinich, A. (Org.), *The Philosophy of Language*. (Original publicado em 1958). Oxford: Oxford University Press.
- Searle, J. (1996b). What is a Speech Act? Em: Martinich, A. (Org.), *The Philosophy of*

Language. (Original publicado em 1965).
Oxford: Oxford University Press.

Searle, J. (1998). *O Mistério da Consciência*.
São Paulo: Paz e Terra.

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. *Am.
Psychologist*, 48, 878–885.

Sternberg, R. (1992). Too young to die –
Let's not bury psychology alive: a response to
Gardner. *New Ideas Psychol.*, 10, 195-205.

Sternberg, R. (2000). *Psicologia Cognitiva*.
Porto Alegre: Artmed.