

Validade científica das teorias de Newell e Simon e de John Searle. Será pertinente a continuação do debate?

Célia Sales
Paulo Santos
Psicólogos

Resumo. – O presente artigo pretende proceder à discussão crítica sobre a validade científica de duas teorias que debatem a possibilidade de Homem e Máquina partilharem as características básicas do pensamento. No centro do debate estão, por um lado, os teóricos da Inteligência Artificial A. Newell e H. Simon e, por outro, o filósofo J. Searle. Separam-nos a hipótese proposta por Newell-Simon – a manipulação simbólica é condição suficiente para o comportamento inteligente – recusada por Searle com a argumentação de que a intencionalidade (característica intrínseca do cérebro humano) é condição necessária para a inteligência, negando que esta propriedade esteja presente em sistemas não biológicos. A comparação das duas teorias foi realizada com base nos critérios de validade científica propostos por Karl Popper. A principal conclusão é a de que a teoria de J. Searle se revela não científica.

O presente trabalho pretende proceder à avaliação de um debate que tem vindo a prolongar-se entre John

Searle e Newell-Simon*. Opõe estas duas facções a possibilidade de considerar tanto o Homem como a Máquina sistemas físicos manipuladores de símbolos, cuja universalidade e capacidade de atingir os objectivos a que se propõem permitem classificá-los como sistemas inteligentes (ou, melhor, sistemas capazes de acção inteligente).

Esta temática remete-nos para velhas questões sobre a natureza da inteligência, do pensamento e da própria mente. Desde a Antiguidade que a resposta a estas questões tem sido uma das preocupações mais ou menos constantes da Filosofia. A Psicologia Cognitiva, interessada na natureza dos processos mentais, vem, já no nosso século, complementar a abordagem filosófica inspirando-se em problemas por esta levantados, operacionalizando-os e procedendo à sua verificação experimental. A Inteligência Artificial (AI), durante algum tempo vista como a base da metáfora computacional ilustrativa dos processos cognitivos humanos, reclama, com a hipótese de Newell-Simon (1980), um papel muito mais central e determinante no processo de descoberta da natureza da mente. Estabelece a manipulação de símbolos como critério para a Inteligência e, existindo

* Agradecemos ao Prof. Leonel Garcia Marques a orientação e comentários durante a realização deste trabalho.

tal competência tanto no Homem como na Máquina (nomeadamente no computador digital), é possível propor uma teoria científica da mente cujo principal objectivo é discriminar as características peculiares da manipulação simbólica humana, partindo de um sistema físico igualmente simbólico (o computador digital).

Esta controvérsia tem vindo a arrastar-se há mais de dez anos. Várias posições foram entretanto tomadas a favor de cada uma destas opções; aparecendo de uma maneira geral os filósofos a favor de Searle e os psicólogos e cientistas das ciências do artificial a favor de Newell-Simon. Não é realmente de estranhar a confusão que deve causar aos filósofos a ideia de que, de alguma maneira, o Homem possa partilhar com a Máquina uma qualidade que desde sempre nos diferenciou dos outros seres vivos: a acção inteligente. E não é também de estranhar que os psicólogos (nomeadamente os adeptos do projecto da *Ciência Cognitiva*) pretendam pelo menos explorar a revolucionária e prometedora hipótese da IA.

Interessa-nos avaliar o impacto deste debate, colocando para tal as seguintes questões: Até que ponto tem sido útil¹ para a *Ciência Cognitiva* o debate Searle *vs* Newell-Simon? Até que ponto poderá vir a ser útil para

a *Ciência Cognitiva* a continuação desse mesmo debate?

Começamos por uma breve caracterização e evolução da ciência cognitiva e da inteligência artificial. De seguida, expomos as teorias em confronto, para depois as avaliarmos à luz das ideias de Popper. Finalmente, respondemos às questões colocadas.

1. *A inteligência artificial como ciência cognitiva.*

Ao longo da História, por umas razões ou por outras, a abordagem cognitiva que aspirava aceder à natureza da mente foi sofrendo revezes, nomeadamente com o advento das teses behavioristas que negavam por completo tal empreendimento. Assim, entidades como estados mentais ou ideias estavam indubitavelmente arredadas da realidade observável, posição que era defendida pelas escolas filosóficas identificadas com o positivismo, fisicalismo e o verificacionismo.

Todavia, o contexto de guerra (início dos anos 40) trouxe desenvolvimentos tecnológicos, dos quais o computador se tornou realidade e depressa se estabeleceram analogias entre os mecanismos bélicos (auto-regulados e autocorrectores) e os processos homeostáticos do sistema nervoso central; estava assim relan-

çado o interesse pelos processos internos e pela abordagem cognitiva. Questões várias se levantaram acerca da natureza desta analogia, nomeadamente quanto à possibilidade de Homem e Máquina partilharem as características básicas do pensamento.

Esta discussão acarretava, entre outros, problemas de objectividade associados à divergência entre os autores sobre o que *é pensar*. Dada a antiguidade do conceito e as várias definições que a Psicologia possibilita, corria-se o risco de todos falarem de entidades diferentes, revelando-se estéril qualquer discussão. O desafio proposto por Turing, em 1950, veio proporcionar uma abordagem objectiva da analogia Homem-Máquina, na medida em que sugeria um critério objectivo para determinar se um computador podia ou não pensar: se um computador estivesse programado por forma a que um observador não pudesse distinguir as suas respostas das de um ser humano, então poder-se-ia afirmar que o computador pensava (Turing, 1963, cit. in Gardner, 1985).

Os dados estavam lançados e «exigia-se» agora uma descrição dos processos cognitivos que permitissem a programação de uma máquina que operasse de modo semelhante ao homem, produzindo o mesmo *output*.

É neste contexto que surge o termo *inteligência artificial* (introduzido por Newell, Simon, McCarthy e Minsky), para denominar um novo campo de estudo, cujo objectivo inicial era o de desenvolver programas que resolvessem problemas, reconhecessem padrões, que pudessem resolver situações de jogo e cuja racionalidade fosse lógica.

Para estes autores, «toda a inteligência envolve o uso e manipulação de vários sistemas simbólicos, tais como aqueles usados na matemática e na lógica» (Gardner, 1985, p. 149), pelo que a equivalência funcional entre o cérebro e o computador não implica qualquer equivalência estrutural ao nível anatómico. Consideram que o importante na resolução de problemas está intimamente ligado à estrutura das operações que permitem chegar à(s) solução(ões), *software*, mais do que aos componentes que permitem essa resolução, *hardware*. «A nossa teoria é uma teoria de processamento de informação que envolve resolução de problemas, e não uma teoria de mecanismos neuronais ou electrónicos para processar a informação» (Newell, Shaw e Simon, 1964, cit. in Gardner, 1985, p. 148).

Aquilo que permitirá acções inteligentes está subjacente ao conceito de *physical symbol system* e constitui a tese central da teoria de Simon e Newell. A partir da crença que os

computadores devem permitir processos semelhantes ao pensamento humano, o objectivo é descrever aqueles aspectos do comportamento de tal modo que possam ser simulados por uma máquina.

A IA, até ao presente, tem contribuído para o desenvolvimento do estudo da mente e, portanto, para o aprofundamento dos mecanismos da cognição humana.

Paralelamente, outros ramos da ciência se debruçaram sobre esta temática, contribuindo também para o enriquecimento das ciências cognitivas. Por exemplo, a Neurociência aprofunda os conhecimentos sobre a natureza neurofisiológica do cérebro humano, adoptando um nível de análise onde a abordagem representacional ou computacional é ignorada; a Antropologia também ignora estas abordagens, dando relevo a factores histórico-culturais para a explicação do funcionamento da mente humana. É neste contexto que nos anos 70 alguns autores propõem que, para se obter uma compreensão o mais completa possível da nossa cognição, é necessário criar pontes de ligação entre áreas de estudo complementares, nomeadamente com a Neurociência e a Psicologia (a um nível micro) e com a Antropologia e outros estudos culturais (ao nível macro).

Surge desta forma uma proposta interdisciplinar que procura desen-

volver linhas de comunicação entre diferentes, mas convergentes áreas das ciências cognitivas, a saber, Filosofia, Psicologia, IA, Linguística, Antropologia e Neurociência (*Hexágono Cognitivo*, proposto pela Sloan Foundation, 1978, cit. in Gardner, 1985, p. 37). Começa pois a ser discutido o termo *Ciência Cognitiva* e, na tentativa de estabelecer o seu alcance, Gardner (1985, p. 6) define-a como uma ciência empírica «que procura responder a questões epistemológicas, em particular aquelas que se preocupam com a natureza do conhecimento, seus componentes, fontes e desenvolvimento. Embora o termo *Ciência Cognitiva* se estenda a todas as formas de conhecimento, seja ele animado ou não, humano ou não, eu aplico o termo principalmente para explicar o conhecimento humano». Debruçando-se sobre as questões levantadas pela Filosofia ao longo de séculos, a *Ciência Cognitiva* considera indispensável centrar-se ao nível de análise das representações mentais, envolvendo entidades como símbolos, regras e imagens e explorando as formas como essas entidades se relacionam e transformam, constituindo o espaço mediador entre *input* e *output*. Defende que este nível das representações não só é legítimo, como necessário para explicar cientificamente a variedade do pensamento, acção e

comportamento humano.

A IA, ao ser usada para simular os processos cognitivos ao nível simbólico/representacional, tem sido uma peça fundamental para a descoberta do funcionamento da mente humana. É neste contexto que se enquadra a teoria de Newell-Simon, propondo, como veremos adiante, uma abordagem científica do tema, reformulando a teoria geral da mente.

2. *A teoria de Newell-Simon. Noção de «sistema físico de símbolos».*

A Ciência Cognitiva pretende compreender a natureza da mente e da inteligência humana; e como foi anteriormente descrito, a IA é uma das várias disciplinas que tem empenhado os seus esforços nesse empenhamento multidisciplinar.

Newell (1980) considera que, neste contexto, a maior contribuição da IA não tem sido o seu uso como instrumento de trabalho (para ajudar a testar modelos do funcionamento cognitivo), mas sim a introdução da noção de *physical symbol systems* (que traduzimos como «sistemas físicos simbólicos»).

Esta noção é importante porque, na sua opinião, constitui uma base para o desenvolvimento de uma teoria científica da mente. Esse carácter científico advém da objectividade com que se pode desenvolver esse

estudo se se seguir a proposta que passamos a explicitar.

2.1. A proposta de Newell para a descoberta científica da natureza da mente.

Dada a complexidade que desde sempre se atribuiu à natureza da mente, Newell propõe que o seu processo de descoberta seja gradual, partindo de pequenas parcelas e chegando ao geral por aproximações:

- 1) Identificar as características da mente (*constraints on mind*) e continuamente tentar descobrir outras novas; Newell (1980, p. 139) aponta treze destas características, salientando que muitas outras poderiam ser incluídas na lista²;
- 2) Aceder à natureza geral da mente através do conhecimento de todas essas características, isto é, compreender tudo o que for possível sobre todos os aspectos nelas existentes;
- 3) Dado que não é possível que uma disciplina (ou dentro da mesma disciplina, um só estudo) consiga abordar este problema, satisfazendo/explicando simultaneamente todas essas características, é desejável que cada um deles, partindo das suas diferentes abordagens e argumentos, incida apenas numa parte.

Constituem-se assim vias convergentes através das quais nos podemos aproximar da verdadeira natureza da mente. Esta é, pois, uma proposta para decompor o problema geral da descoberta da mente;

- 4) É no entanto aconselhável que, dentro da mesma disciplina, a hipótese/teoria considerada satisfaça o maior número de restrições. Por isso deve-se começar por modelos ou sistemas que satisfaçam uma ou duas dessas características e que sejam «generativos», isto é, que pareçam passíveis de gerar modelos que satisfaçam outras características (e continuarem ainda assim generativos).

A IA segue esta proposta, construindo um sistema físico que satisfaz, à partida, duas das características da mente: a universalidade e o comportamento simbólico (capacidade de abstracção). Pretende conseguir a maior aproximação possível deste sistema físico à mente humana, ao inculir-lhe uma a uma as características da mente. Assim, vai por exemplo tentar determinar as condições para que esse sistema físico universal e manipulador de símbolos obedeça também à racionalidade. E, depois de conseguir construir um sistema físico universal manipulador de símbolos e racional, irá

tentar determinar as condições em que ele é capaz de satisfazer uma outra característica da mente, como a operação em tempo real... A IA pretende, pois, partir de uma classe geral de sistemas físicos (que partilham entre si o facto de manipularem símbolos e apresentarem universalidade de comportamento), para classes com características cada vez mais específicas, que as aproximam cada vez mais do funcionamento cognitivo humano.

Um sistema físico de símbolos é qualquer máquina que possua as operações básicas necessárias para o processo representacional (ace-der, copiar, ler, escrever e interpretar, do inglês *assign, copy, read, write, e interpret*).

No seu conjunto, estas operações garantem a universalidade do sistema, isto é, garantem que ele seja capaz de realizar (quando bem instruído) qualquer sequência *input-output*.

Esta propriedade é muito importante, pois corresponde à flexibilidade de comportamento desejável num sistema inteligente, de produção de um *output* de acordo com qualquer *input* não previsível à partida³, adaptando-se assim às exigências do meio.

2.2. Hipótese dos sistemas físicos simbólicos.

O que esta hipótese propõe é que este comportamento simbólico (presente no sistema físico descrito) é universal, é o único que existe, incluindo portanto o comportamento simbólico humano⁴.

Newell (1980, p. 170) apresenta formalmente esta hipótese da maneira que passamos a citar: a condição necessária e suficiente para um sistema físico exibir acção inteligente é que seja um sistema físico simbólico.

Condição necessária. Significa que qualquer sistema físico que exiba inteligência geral é um sistema físico simbólico.

Condição suficiente. Significa que qualquer sistema físico simbólico pode ser organizado por forma a exibir acção inteligente geral.

Ação Inteligente Geral. Significa o mesmo que a inteligência que é observada no ser humano: significa, em situações reais, a ocorrência de comportamento apropriado aos fins do sistema e adaptado às exigências do meio (dentro de alguns limites físicos).

3. *A teoria de Searle.*

A posição de Searle face à Inteligência Artificial baseia-se numa dicotomia que ele próprio estabeleceu entre *Weak AI* e *Strong AI* («versão fraca» e «versão forte», respectivamente). Concorde com a versão fraca, isto é, considera importante e legítimo o papel da IA como instrumento de trabalho no estudo da cognição humana, nomeadamente para testar teorias através da simulação. Opõe-se à versão forte, que definiu como a hipótese de que «um computador devidamente programado tem estados cognitivos, e nesse sentido explica a cognição humana» (Searle, 1980, p. 417).

A versão forte de Searle corresponde à hipótese proposta por Newell-Simon, da qual o primeiro discorda, alegando que um programa de computador por si só não é suficiente para explicar o pensamento (a máquina, dadas as suas características intrínsecas, e não o programa, é que o deverá fazer); centra os seus argumentos na questão de que a manipulação simbólica não é suficiente para que uma máquina pense, porque essa manipulação não garante a compreensão e a intencionalidade daquilo que processa. Para provar este ponto dá o seu famoso exemplo do quarto chinês⁵, em que um homem só manipulando símbolos, associando regras de sintaxe que lhe

vão dando, consegue falar chinês sem, no entanto, compreender o que faz (isto é, sem alcançar o significado ou semântica do seu *output*). Neste contexto, Searle apresenta uma teoria que pretende refutar a hipótese de sistemas físicos simbólicos. Segue-se um resumo dessa teoria.

1) Searle considera que a diferença entre mecanismos humanos e não-humanos é a compreensão ou a intencionalidade. Neste sentido, nega que uma máquina possa compreender. A questão da compreensão por parte de mecanismos não-humanos é assumida pelo autor como uma «atribuição metafórica de intencionalidade» (Searle, 1980), na medida em que o sentido em que as máquinas percebem não é o mesmo em que nós (humanos) percebemos. Para Searle a IA deve explicar o especificamente mental, distinguindo-o daquilo que não o é; e isso deve ser feito, não explicando aquilo que é visto pelo investigador (pela observação do *output* do sistema), mas por aquilo que o sistema tem de intrínseco.

2) Assim, a intencionalidade é produto daquilo a que ele chama «poderes causais do cérebro». Qualquer simulação estrutural, ou de sequências de impulsos ao nível neuronal, não é condição suficiente e, por consequência, não se traduz na habilidade para produzir estados

intencionais; estes não têm nada a ver com programas formais. «A simulação da cognição não produzirá os efeitos da neurobiologia dos estados cognitivos» (Searle, 1990, p. 24). Para conceber a cognição, não basta a simples manipulação de símbolos, ainda que em interação com o meio porque...

3) Os processos computacionais e os *outputs* podem existir sem que haja um estado cognitivo. «O problema em discussão não é acerca de saber como é que eu chego aos estados cognitivos dos outros, mas o que há em mim quando lhes atribuo estados cognitivos» (Searle, 1980, p. 422).

4) Em suma, baseado no argumento de que nós somos uma certa espécie de coisa biológica (física e/ou química), ele conclui que um programa falha na explicação dos processos mentais porque a intencionalidade é fruto de poderes causais específicos dessa coisa biológica; o programa não é produto do computador (máquina), enquanto os estados mentais são produto de operações no cérebro. A manipulação de símbolos só obedece a regras, é apenas sintaxe; não corresponde sequer ao processamento de informação porque a correspondência entre os símbolos processados e entidades reais é feita pelo homem, não pelo computador. De acordo com as próprias palavras de Searle, «Repetindo, o computa-

dor tem sintaxe mas não semântica» (1980, p. 423).

4. *Popper: critérios de discussão crítica de teorias.*

Optámos por avaliar este debate à luz dos critérios propostos por Karl Popper para a discussão crítica dos méritos das teorias científicas. Esta opção prende-se não só com o facto de considerarmos Popper como um autor versátil, que apresenta de forma clara os critérios relevantes que permitem comparar teorias; mas também por nos parecer apropriado para esclarecer algumas questões relativas à teoria de Searle, nomeadamente quanto à possibilidade que esta oferece de (não) ser refutada. Para Popper, o conhecimento científico só evolui desde que haja espírito crítico. A essência deste espírito está patente na ideia central de que a observação só serve para falsificar (razão crítica) e nunca para justificar (razão positiva).

Ao avaliar uma teoria não se deverá usar critérios positivos que a justifiquem e que a definam como melhor ou pior do que a sua adversária. Pelo contrário, há que avaliá-la pela negativa, criticá-la, tentar provar que ela é falsa.

Para escolher entre teorias concorrentes, Popper considera que se deve tentar refutá-las. Aquela que me-

lhor suportar as críticas é a mais válida nesse momento, aproximando-se mais da «verdade».

Não há, pois, teorias verdadeiras. Todas as teorias (tal como todas as críticas) são meras conjecturas: são válidas apenas num determinado momento, ou seja, enquanto não forem provadas falsas. Daí que não se fale em «verdade» mas em «grau de verosimilhança» de uma teoria (situação em que esta se aproxima da verdade). Tal como já foi descrita, essa proximidade é tanto maior quanto maior for a resistência a críticas e testes, o seu «grau de corroboração».

Repare-se que quando consideramos que uma teoria é mais válida do que outra, estamos a acreditar na sua proximidade à verdade. De facto, existe uma «crença», que será tanto mais racional quanto maior o grau de corroboração. Diz Popper: «A racionalidade da ciência e dos seus resultados – e, portanto, da ‘crença’ neles – é inseparável, por essência, do seu progresso, com a discussão actualizada dos méritos relativos de novas teorias; é inseparável do progressivo derrube de teorias» (1987, p. 86).

Baseando o avanço da ciência neste «derrube progressivo de teorias», Popper propõe que a distinção entre ciência e não-ciência reside na refutação, logo, na possibilidade de teste.

4.1. Critérios de comparação.

É exactamente esta possibilidade de teste um dos critérios a ter em conta, na opinião deste autor, numa discussão crítica sobre a validade relativa de duas teorias: «se não dispusermos de meios para submetê-la a testes, não haverá para ela evidência empírica e, portanto, não haverá como proclamá-la científica» (Magee, 1973, p. 50).

Existem, no entanto, outros critérios. O *conteúdo informativo* é um deles (Magee, 1973). Uma teoria deve encerrar um alto conteúdo informativo, isto é, deve ser o mais completa, específica e precisa na descrição da realidade com que está a lidar, e ao mesmo tempo não-*tautológica*. Este critério está intimamente relacionado com o da refutabilidade. Isto porque quanto maior for o conteúdo informativo de uma teoria, maior a sua possibilidade de teste.

Temos igualmente que considerar se as teorias em causa resolvem o problema que é suposto resolverem; se explicam o que é suposto explicarem; se não se limitam a alterar o problema, através, por exemplo, de uma suposição *ad-hoc* não testável (valor explicativo).

A inovação proposta pela teoria face às anteriores é também outro factor a ter em conta. De facto, de nada adianta que ela seja refutável, que

tenha um alto conteúdo informativo e um valor explicativo, se não acrescentar nada de novo àquilo que já se sabe. É neste sentido que se pode considerar uma teoria relevante ou não.

5. *Comparação das duas teorias, de acordo com Popper.*

5.1. Refutabilidade.

Em relação a Searle, não encontramos uma definição explícita daquilo que ele entende por intencionalidade. Quando muito, poderemos inferir que ele parte do princípio de que a intencionalidade é a interpretação/compreensão dos símbolos, ou seja, a atribuição de significado (semântica) à manipulação simbólica. Ao considerar que essa atribuição de significado se deve aos «poderes causais do cérebro», o autor apresenta um enunciado irrefutável: não define o que são «poderes causais do cérebro», não especificando o que confere propriedades semânticas aos símbolos; apenas associa esses poderes a certas características bioquímicas específicas do cérebro. Ao não se apontar concretamente quais são, não é possível testar que elas não existem nem tão-pouco que elas não causam a intencionalidade. Nesta perspectiva, este enunciado não é científico por-

que não existe maneira de mostrá-lo falso.

Encontrámos, na teoria de Searle, um aspecto que pode ser refutado: a sua afirmação de que os computadores só operam ao nível da sintaxe, isto é, não denotam capacidade para atribuir significado (semântica) aos símbolos. No entanto, este aspecto já foi refutado. Já foi concebido um programa (ZBIE, desenvolvido por Siklóssy, 1972, cit. in Simon, 1992) que compreende as frases que lê: o ZBIE representa em diagramas o conteúdo das frases estabelecendo, assim, a relação entre sintaxe e semântica.

A teoria de Newell-Simon parece-nos globalmente refutável na medida em que a hipótese (na versão necessária e na versão suficiente) admite a possibilidade de ser falsificada mediante o recurso a contra-exemplos.

Assim, para refutar a condição necessária bastará encontrar um sistema físico inteligente que não seja um sistema físico simbólico. Para refutar a condição suficiente bastará encontrar um sistema físico simbólico que não exiba acção inteligente geral.

Actualmente existe uma teoria cujo desenvolvimento poderá vir a refutar a condição necessária. Os PTC (Proper Treatment of Connectionism) preconizam a acção inteligente através da manipulação sub-

-simbólica, adoptando um nível de análise (e de explicação dos processos cognitivos) que se situa entre o modelo de redes neuronais e o paradigma simbólico (Smolensky, 1988).

Poderemos dizer que Searle tenta refutar a condição suficiente: ele afirma que a máquina (sistema físico manipulador de símbolos) não exibe acção inteligente (porque não é dotado de intencionalidade). Defende, pois, que a manipulação de símbolos não é suficiente para a caracterização dos estados mentais. Uma outra forma de encarar a crítica de Searle é vê-la como uma divergência ao nível da concepção de inteligência. Ele defende que a adequação de uma resposta ao *input* (e, portanto, ao meio) não é razão suficiente para que a possamos dizer inteligente (contrariando a noção de «acção Inteligente Geral» de Newell-Simon). Considera que é necessário entrar em linha de conta com aspectos intrínsecos, que a um nível neurofisiológico conferem intencionalidade às operações simbólicas. No entanto, não só (como já vimos) esses factores não são identificados, como (tal como iremos ver) o baixo conteúdo informativo acerca das suas supostas condições de acção ou de inibição dificultam, mais uma vez, quaisquer tentativas de refutação.

5.2. Conteúdo informativo.

A teoria de Newell-Simon encerra um alto conteúdo informativo, dado que é apresentada de forma minuciosa, caracterizando todos os passos e pressupostos em que se baseia a formulação da hipótese. A riqueza de conteúdo informativo torna igualmente rica a possibilidade de refutação ao nível desses pormenores focados.

Sendo a teoria de Searle considerada metafísica de acordo com o critério de refutabilidade popperiano, isso não significa que ela não possa ser eventualmente aproveitada e refutada.

A possibilidade de refutação de uma teoria é que a torna empírica, ou seja, quando existe uma classe de proposições-teste empíricas que a contradizem (Weingartner, cit. in Lorenz e Popper, 1983). É pelo facto de a teoria ser falsificável que se lhe atribui conteúdo. Deste modo, somos levados a concluir que a teoria de Searle é desprovida de conteúdo. Senão, o que nos diria ele sobre as condições em que os poderes causais actuam?

Vejamos o exemplo do quarto chinês. Aí o indivíduo está sujeito a duas condições experimentais: uma em que se limita a obedecer a um conjunto de regras escritas; outra em que responde a perguntas que lhe são formuladas em inglês. Em

ambas as condições, o sujeito, possuidor de um cérebro, é potencialmente capaz de usar os poderes causais. Na opinião de Searle, isso não acontece na primeira condição porque o homem está simplesmente a manipular regras simbólicas sem significado. Só na segunda condição os poderes causais actuam, e é por isso que há intencionalidade. Como é que Searle garante que na primeira condição os poderes causais não actuaram? Como refuta ele a afirmação «Na primeira condição, os poderes causais actuaram», ou a afirmação «Os poderes causais do indivíduo agiram previamente no sentido de não agirem em contexto experimental»?

Para além de não especificar estes (e quase todos os) pormenores em que baseia a sua teoria, Searle dá explicações tautológicas: porque há poderes causais, então há intencionalidade... e esta é fruto dos poderes causais. O conteúdo informativo da sua teoria é, pois, muito baixo.

5.3. Valor explicativo.

O valor explicativo da teoria de Newell-Simon parece ser razoável. O que ela pretende mostrar é que o Homem partilha com outras entidades não biológicas capacidades simbólicas universais, sendo no entanto uma subclasse particular

dessa classe de sistemas manipuladores de símbolos. Essa especificidade, no entanto, não influencia o facto de a sua acção inteligente advir dessa capacidade representacional. Newell e Simon pretendem mostrar a independência da manipulação de símbolos do tipo de implementação que a suporta; e querem desenvolver mecanismos que se aproximem cada vez mais da forma peculiar como o Homem faz essa manipulação. De facto, estes autores explicam o que pretendem explicar quando identificam *a*) as operações universais suficientes para que qualquer máquina possa manipular símbolos; e *b*) quando apontam possíveis soluções para a construção de sistemas que a pouco e pouco abranjam as características da representação mental humana.

Se pensarmos no que Searle pretende explicar com a sua teoria e aquilo que de facto é explicado, verificamos que existem lacunas. Tal como já foi mostrado, as suas pretensões de que factores neurofisiológicos estão envolvidos na determinação de intencionalidade não são esclarecidas, dado o reduzido conteúdo informativo das suas formulações.

5.4. Inovação.

A hipótese de sistemas físicos simbólicos é altamente inovadora, no

sentido em que propõe uma nova forma de perspectivar a investigação em psicologia cognitiva. Ela propõe uma abordagem científica de um assunto que desde sempre se tem revestido de subjectividade e impasses desta derivados.

Já Searle, e de acordo com uma crítica de Fodor (Searle, 1980), parece querer contrariar a versão forte com base em argumentos desde há muito usados por filósofos: «Esta situação – assumir a existência de uma relação causal sem saber exactamente qual ela é – é inteiramente familiar em filosofia» (Fodor, cit. in Searle, 1980, p. 431). Tendo em conta o que já foi dito, a sua proposta não parece trazer nada de novo.

6. Conclusão.

Passamos a responder à primeira questão, «Até que ponto tem sido útil para a *Ciência Cognitiva* o debate Searle vs Newell-Simon?». Pelo facto de a teoria de Searle levantar questões filosóficas acerca da natureza dos estados mentais, pode ter constituído um estímulo para o desenvolvimento de pesquisas no âmbito da *Ciência Cognitiva*.

Segunda questão, «Até que ponto poderá vir a ser útil a continuação desse mesmo debate?». Não vemos utilidade na continuação do debate, porque *a*) Searle não parece ter le-

vado em conta a evolução do conhecimento das ciências cognitivas na última década. Ele não procurou integrar as inovações em novas questões filosóficas que reflectam as grandes preocupações actuais nesta área; *b*) este autor rejeita a hipótese de Newell-Simon, contrapondo-lhe a sua teoria, a qual, quando analisada à luz de Popper, se revela globalmente não-científica, com baixo conteúdo informativo, sem valor explicativo e não inovadora. Portanto, estando estas duas teorias em competição, a de Newell-Simon é a mais válida; *c*) a teoria de Searle já foi falsificada no único ponto refutável que apresenta.

ças entre máquinas universais mas, embora importantes, essas diferenças não afectam a natureza dos símbolos por elas partilhadas. Tal como H. Simon (1969) refere, entende-se o Homem e a Máquina como sistemas simbólicos, como entidades que processam, transformam, elaboram e manipulam símbolos de diversas maneiras.

⁵ O exemplo do quarto chinês consiste na hipotética situação de um indivíduo isolado num quarto, que não sabe falar chinês e a quem são fornecidos símbolos gráficos dessa língua, assim como instruções – regras de sintaxe – para a construção de frases. Face a instruções correctas e completas, ele consegue construir frases em chinês sem, no entanto, perceber o seu significado. Searle defende que o processo ilustrado neste exemplo é semelhante à manipulação de símbolos pela Máquina.

¹ Útil, no sentido de ajudar a esclarecer a natureza dos processos mentais.

² Exemplos: universalidade, operação em tempo real, comportamento racional, uso de linguagem, uso de um vasto conhecimento sobre o meio, etc.

³ Esta universalidade é relativa. Não há nenhum sistema capaz de realizar todo o tipo de respostas. O próprio Homem tem limitações físicas que restringem o seu comportamento. Assim, a noção de universalidade é relativa a uma classe; uma máquina que pertence a uma classe é dita universal se conseguir comportar-se como qualquer outra máquina dessa classe.

⁴ Newell não pretende que todas as máquinas universais sejam iguais, ou (nesta ordem de raciocínio) que o Homem seja igual à Máquina. Existem diferen-

Referências

- Foundation, S. (1978), *Cognitive science 1978*, Report of the State of the Art Committee, Nova Iorque.
- Gardner, H. (1985), *The mind's new science. A history of the cognitive revolution*, Basic Books, Nova Iorque.
- Magee, B. (1973), *As Ideias de Popper*, Cultrix, São Paulo.
- Newell, A., Shaw, J. C., e Simon, H. A. (1964), «Elements of a theory of human problem solving», in R. J. C. Harper, C. C. Anderson, C. M. Christensen, e S. M. Hunka (eds.), *The cognitive processes: Readings*, Englewood Cliffs, N.J.: Prentice Hall. Reprinted da *Psychological Review*, 65 (1958).

- Newell, A. (1980), «Physical Symbol Systems», in *Cognitive Science*, 4, pp. 135-83.
- Popper, K. (1987), *O realismo e o objectivo da ciência*, vol 1º do pós-escrito à lógica da descoberta científica, Dom Quixote, Lisboa.
- Popper, K.; e Lorenz, K. (1983), *O futuro está aberto*, Fragmentos, Lisboa.
- Searle, J. (1980), «Minds, brains and programs», in *The Behavioral and Brain Sciences*, 3, pp. 417-57.
- (1990), «Artificial intelligence: A debate», in *Scientific American*, 262 (1), pp. 19-25.
- Siklóssy, L. (1972), «Natural language learning by computer», in H. A. Simon e L. Siklóssy (ed.), *Representation and meaning*. Englewood Cliffs, N. J.: Prentice-Hall.
- Simon, H. (1969), *As Ciências do Artificial*, Dom Quixote, Lisboa.
- (1992), «What is an 'explanation' of behavior?», in *Psychological Science*, 3 (3), pp. 150-61.
- Smolensky, Paul (1988), «On the proper treatment of connectionism», in *Behavioral and Brain Sciences*, 11, pp. 1-74.
- Turing, A. M. (1963), «Computing machinery and intelligence», in E. A. Feigenbaum e J. Feldman (eds.), *Computers and thought*, Nova Iorque, Mc Graw-Hill. Edição original, 1950.
- A. Newell e H. Simon state that symbolic manipulation is a sufficient condition for intelligent behavior. J. Searle denies that possibility arguing that intencionality – an intrinsic characteristic of the human brain – is a necessary condition for intelligence, and cannot be shared by non-biological systems. We analyse both theories according to Karl Popper's scientific validation criteria. The main conclusion is that Searle's theory is not a scientific one.

Abstract. – This paper is a critical discussion on the scientific validity of two antagonist theories which debate the possibility of Man and Machine sharing the same basic characteristics of thought.