Psicologia (1983) IV, 1 e 2: 77-80

Programas, estados cognitivos e explicação psicológica

MARIA LUÍSA FIGUEIRA *

A ciência cognitiva não é rigorosamente um corpo teórico de conhecimentos mas um conjunto de projectos, supostamente transdisciplinares, envolvendo várias disciplinas das ciências do Homem e disciplinas formalizantes, cujo uso de linguagens abstractas, de rápida evolução tecnológica, veio produzir um progresso, apenas aparente, para a compreensão científica da natureza do acto de conhecer. Estes pressupostos de princípio são refutados pelos que afirmam ser a ciência cognitiva, pelo contrário, uma verdadeira teoria psicológica explicativa, cuja base experimental de confirmação se apoia na «Inteligência Artificial». Para os defensores desta tese, os complexos programas de computador que foram implementados, seriam um modelo do funcionamento psíquico ou mental – existindo uma identificação, em sentido literal, entre os estados e processos do programa e os estados e processos mentais. Os computadores, se adequadamente programados, teriam então *estados cognitivos* e os programas seriam teorias psicológicas com poder explicativo. De acordo com estes postulados (que foram designados por «Inteligência Artificial Forte») as máquinas programadas teriam um poder causal que derivaria não da substância ou material de que são constituídas, mas do desenho e dos programas que nelas «corriam». Do mesmo modo os estados cognitivos existiam no cérebro mas não tinham uma relação directa com a «substância» neuronal da sua constituição. Nesta perspectiva pressupõe-se a irrelevância dos mecanismos neurofisiológicos ou neuropsicológicos do Sistema Nervoso Central, optando-se pelos mecanismos abstractos ou puramente formais; pelo menos considera-se como implícita a irredutibilidade (ou uma conexão não essencial) entre os dois níveis de discurso ou os dois universos em sentido de Popper. O critério de decisão sobre o carácter «mental» ou «cognitivo» do programa de computador é em geral considerado o teste desenhado por Turing, isto é, se o sistema pode convencer um perito competente(que lhe é exterior do ponto de vista espacial e mental) que tem estados cognitivos, então ele possui de facto esses estados. Para os defensores doutra posição mais restritiva (designada por «Inteligência Artificial Fraca») o principal valor do computador no estudo dos processos psicológicos em particular dos cognitivos, seria o de um poderoso instrumento que permitiria aos investigadores a formulação e o teste da hipóteses

* Dpt. de Psiquiatria, Faculdade de Medicina da Universidade de Lisboa.

duma forma mais rigorosa do que nunca. O poder explicativo não estaria contido no próprio programa ou na máquina programada mas seria por inferência indirecta o inerente à teoria psicológica a testar.

As reflexões críticas sobre estas afirmações introdutórias poderiam ser inúmeras e tem constituído matéria de discussão de muitos autores entre os quais poderíamos citar H. L. Dreyfus, H. Putnam, D. C. Dennett, J. Searl, J. A. Fodor e J. Haugeland entre outros (ver bibliografia). O carácter exterior ou interior destas críticas em relação à própria disciplina – Inteligência Artificial (I.A.) – ou o seu maior ou menor radicalismo variam consideravelmente, no entanto como paradigma e porque tem em nossa opinião razoável consistência interna e pertinência face aos problemas epistemológicos levantados, iremos considerar a crítica formulada em vários artigos por John Searl e condensada, na sua quase totalidade, no trabalho *Minds, Brains and Programs* (1980).

A posição em que Searl se coloca para efectuar as teses da I.A. «forte» é a de tentar demonstrar através de contra-exemplos que existem diferenças tão essenciais entre os programas e os processos ou estados cognitivos que não é possível fazer uma identificação entre ambos. Num dos seus contra-exemplos mais conhecidos que é o da compreensão de uma história em chinês, pretende contestar, de modo exemplar, o trabalho de Roger Schank e dos seus colegas de Yale que construíram um programa em que se pretende simular a capacidade humana de compreender histórias. No entanto os mesmos argumentos, segundo afirma o próprio Searl, se poderiam aplicar ao programa de Winograd (1972) SHRDLU, de Weizenbaum (1965) ELIZA e dum modo geral a todas as simulações dos processos mentais humanos pela máquina de Turing. O programa de R. Schank (1977) tem a finalidade de simular a capacidade humana de compreender histórias, considerando-se que se for fornecida ao computador uma determinada história e se ele for «capaz» de imprimir respostas a perguntas que lhe forem colocadas, dum modo sequencial, a propósito da história mesmo referentes a informação que apenas está implícita, tal como poderia fazer um ser humano numa situação semelhante então considera-se que: a máquina *compreendeu* literalmente a história hipótese 1) e que o programa *explica* a capacidade humana de compreender a história e de responder a perguntas sobre ela (hipótese 2). Por exemplo, suponhamos a seguinte história: «Um homem entra num restaurante e pede um *hamburguer*. Quando o *hamburguer* chega verifica que está queimado. Quando o homem sai zangado do restaurante não paga a conta ou deixa gorgeta.» Neste momento se perguntarmos «comeu o homem o *hamburguer*?» a presumível resposta será «não». Do mesmo modo se a história for a seguinte: «Um homem entra num restaurante e pede um *hamburguer*. Quando o *hamburguer* chega ele fica muito satisfeito com ele. Quando sai do restaurante dá ao criado uma grande gorgeta antes de pagar a conta.» À questão «comeu o homem o *hamburguer*?», a resposta será naturalmente «sim». A máquina programada por Schank pode responder a questões semelhantes se tiver uma «representação» do tipo de informação que os seres humanos têm acerca dos restaurantes e que lhes permitem responder a questões como as que citámos. As respostas que a máquina imprime são semelhantes às que daria um ser humano em circunstâncias semelhantes. A conclusão tirada pelos defensores da I.A. «forte» não é apenas que esta sequência de perguntas-respostas simula a capacidade humana, mas que a máquina compreende a história e que o programa tem um poder explicativo dessa capacidade humana, conforme referimos nas duas hipóteses 1 e 2.

Estes pressupostos são fortemente contestados por Searl e o contra-exeplo que utiliza baseia-se na hipótese de que podemos testar esta teoria se perguntar-mos a nós mesmos como é que seria se a nossa própria mente trabalhasse segundo os princípios que a teoria afirma.

Suponhamos que a um sujeito S colocado num quarto fechado era fornecido um texto em chinês. Se este não conhecer a língua chinesa (escrita ou falada) não será capaz de reconhecer a escrita chinesa dum modo distinto por exemplo da escrita japonesa. Em seguida é-lhe dado um segundo texto em escrita chinesa e um conjunto de regras que permitem correlacionar o primeiro texto com o segundo. Estas regras seriam escritas numa língua, por exemplo inglês, que o sujeito S compreende perfeitamente e conhece tão bem como qualquer outro sujeito de naturalidade inglesa. O que as regras possibilitariam era que o sujeito pudesse fazer a correlação entre um conjunto de símbolos formais – a palavra formal é utilizada por Searl no sentido em que os símbolos podem ser identificados inteiramente pela sua forma. Suponhamos que

também se fornece ao sujeito um terceiro texto com símbolos chineses em conjunto com algumas instruções ou regras, de novo em inglês, que agora lhe permite correlacionar elementos desses terceiro texto com os dois primeiros. Estas regras possibilitam que o sujeito «devolva» certos símbolos com certas formas (chinesas) em resposta a certas formas contidas no terceiro texto. Supondo-se desconhecidas para o sujeito, as pessoas que lhe estão a transmitir os símbolos podem designar o primeiro texto por «argumento», o segundo texto por «história», o terceiro texto por «perguntas», os símbolos devolvidos por S por «respostas às perguntas» e ao conjunto das regras em inglês por «programa». Neste momento temos o contra-exemplo completamente construído e se o sujeito S estiver de tal modo treinado nas instruções para manipular os símbolos em chinês e se os programadores forem tão eficazes a escreverem o programa, dum ponto de vista externo – isto é do ponto de vista de alguém que estaria fora do quarto onde S estaria fechado – as respostas do sujeito S às questões podem ser *indistinguíveis* das que daria qualquer outro sujeito de naturalidade chinesa. Nenhum observador externo poderia portanto afirmar que o sujeito S não era capaz de falar uma palavra de chinês. Podemos ainda ir mais longe e afirmar que neste ponto da situação construída, as respostas do sujeito S eram também indistinguíveis – do ponto de vista da compreensão da língua – de uma outra situação que poderíamos imaginar: as respostas dum sujeito a quem as histórias eram fornecidas em inglês (língua que ele conhecia bem), as perguntas acerca da história eram escritas em inglês e as respostas do sujeito eram também dadas em inglês. Mais uma vez um observador externo ao ler as respostas quer em relação às perguntas em inglês quer em relação às perguntas em chinês iria concluir serem ambas igualmente boas. No entanto, no caso «chinês» o sujeito S produz as respostas pela *manipulação de símbolos formais não interpretados* e comporta-se como um computador, isto é, executa operações computacionais sobre símbolos formais não identificados. Tanto o sujeito como a máquina não compreendem uma palavra das histórias chinesas embora os seus «inputs» e «outputs» sejam indistinguíveis dos de seres (humanos) que fossem naturais da China. Deste modo, e pela análise desta conclusão, fica segundo Searl refutada a hipótese que designámos por hipótese 1. No que respeita à possibilidade do programa ser explicativo em relação à capacidade humana de compreender histórias – hipótese 2 – a sua contestação deduz-se naturalmente da primeira. Isto é, se o programa e o computador não fornecem condições suficientes de compreensão também não podem ter um poder explicativo. De notar que o termo compreensão – que implica a existência de estados mentais intencionais com um determinado valor de verdade é, para fins de discussão, utilizado apenas em relação ao aspecto da existência ou posse desses estados. Podemos então afirmar que o poder explicativo ou causal não existe no programa – que é um modelo formal – dado que *as suas propriedades formais não possuem em si mesmas qualquer intencionalidade* excepto o poder de produzir ou causar o estado seguinte do seu próprio formalismo, quando o programa está «correndo» na máquina. As manipulações dos símbolos formais são *destituídas de significado* dado que os símbolos não possuem propriedades simbólicas, possuem uma sintaxe mas não uma semântica. A intencionalidade e o significado que os computadores parecem possuir está, segundo Searl, apenas na mente dos que elaboraram o programa, dos que fornecem o «input» e interpretam o «output». O carácter puramente formal do programa por si só exclusivo de qualquer intencionalidade que só poderá ser defenida em termos do seu conteúdo. Tal como um sistema linguístico as expressões sintáticas podem possuir um extenso grau de variedade, sendo o seu conteúdo semântico o que lhes confere significado e direccionalidade.

As objecções postas a esta posição de Searl pelos teóricos da I.A. têm sido extensas e cobrem vários campos de crítica, o próprio Searle as analisa e refuta no artigo que citamos e embora interessantes são demasiado extensas para este curto trabalho. Em relação à argumentação de Searl no caso «chinês» quereríamos colocar, nós mesmos, duas questões preliminares é evidente que o reconhecimento dos símbolos da escrita apenas pela sua forma não implica a compreensão do seu significado. No entanto, esse mesmo reconhecimento só é possível se existirem *unidades formais mínimas e invariantes* duma dada língua que tem uma correspondência com os seus fonemas. Nesse sentido podemos considerar que existe um nível semântico mínimo subjacente à obtenção das respostas dadas pelo computador. O problema epistemológico existe se se pretende fazer uma generali-

zação desse nível semântico mínimo e restrito para o nível e universo semântico da compreensão humana de uma língua, e assim identificar os dois estados.

Se o contra-exemplo de Searl nos permite uma intuição do funcionamento dum programa de simulação do tipo «compreensão de histórias», algumas críticas epistemológicas mais gerais poderiam ser referidas. Assim se considerarmos que os estados ditos mentais não são apenas cognitivos mas, por exemplo, afectivos as objecções de fundo às teses da I.A. «forte» ficam mais claras sobretudo se tentarmos uma concretização: por exemplo se obtivermos no «output» do computador a seguinte resposta «estou triste», ninguém se lembraria de afirmar que a máquina está verdadeiramente triste ou que o seu estado nesse momento é identificável com o estado humano de tristeza, no entanto se a resposta do computador for «penso que o homem comeu o *hamburguer*» porque afirmar que esse é um estado com realidade cognitiva e que de facto a máquina pensa? Se no primeiro caso nos é fácil separar o nível formal do semântico no segundo ficamos ancorados numa forma de realismo que consiste em atribuir a qualidade do real ao conhecimento de algumas das suas propriedades. Em relação a aspectos filosóficos gerais das teses da I.A. «forte» é ainda Searl quem afirma haver nessa teoria dois tipos de resíduos conceptuais, qualquer dos dois correspondendo a posições teóricas fortemente repudiadas por esses mesmos teóricos: 1 – um operacionalismo ou «behaviourismo» residual contido no carácter de realidade concedido a semelhanças que são puramente empíricas e obtidas dum modo operacional; 2 – uma forma residual de dualismo (variedade do dualismo cartesiano clássico) no projecto de reproduzir e explicar o mental através do desenho de programas na suposição de que a mente ou o espírito estão não só conceptualmente mas empiricamente separadas do cérebro tal como é o próprio programa da substância da máquina. A inexistência de qualquer conexão intrínseca entre o que há de especificamente mental (nos estados cognitivos humanos e no programa) e as reais propriedades materiais (do cérebro ou do computador) que se admite como pressuposto esconde o verdadeiro dualismo das diferenças de «substância».

Estaremos de facto numa época de concepção mitológica do computador? Ou estaremos a ser simplesmente presas de obstáculos epistemológicos ao conhecimento científico: tentações generalizantes, substancialismo de qualidades de superfície, identificações globalizantes de níveis diferentes de realidade? É que aquilo que está em causa, não é que nós seres psicológicos, sendo máquinas pensantes possamos conceber a existência de outras máquinas que possam pensar, nem mesmo que os computadores digitais possam ser essas máquinas a um certo nível do «pensamento». O que está verdadeiramente em causa é que possamos afirmar que uma máquina pelo simples facto de ser um computador com um programa correcto e eficaz possua a condição suficiente para ter inteligência.

REFERÊNCIAS

DENNETT, D. C. (1971) – «Intentional Systems», *The Journal of Philosophy*, 68, 87-106, reimpresso em *Brainstorms*, Bradford Books, M. I. T., 1978.

DREYFUS, H. L. (1979) – *What Computers Can't Do*, New York: Harper and Row, 1979.

FODOR, J. A. (1980) – «Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology», *The Behavioral and Brain Sciences*, 3, 63-73. Reimpresso em *Mind Design*, ed. J. Haugeland. M.I.T. Press, 1981.

HAUGELAND, J. (1978) – «The Nature and Plausibility of Cognitivism», *The Behavioral and Brain Sciences*, 1, 215-226. Reimpresso em *Mind Design*, ed. J. Haugeland, M.I.T. Press, 1981.

SCHANK, R. (1975) – *Conceptual Information Processing*, North-Holland Publishing Company, Amsterdam.

SEARL, J. R. (1980) – «Minds, Brains and Programs», *The Behavioral and Brain Sciences*, 3, 417-424. Reimpresso em *The Mind's I*, ed. D. Hofstadter, D. C. Dennett, Basic Books, New York, 1981.

SEARL, J. R. (1982) – «The Myth of Computer», *The New York Review of Books*, Vol. XXIX, 7, 3-6.