

Daniel Andler*

O projecto cognitivista

Os fenómenos cognitivos são abordados de diversos modos. Entre as disciplinas que deles fazem objecto de estudo temos de mencionar as neurociências, a psicofísica, a psicologia, a linguística, a informática, a mecânica, a robótica e também a lógica, a filosofia e mesmo, desde há pouco tempo, a física dos meios desordenados. Mas o centro do terreno é ocupado, desde há trinta anos, por duas especialidades que formam uma espécie de par. Trata-se da inteligência artificial e da psicologia cognitiva (por vezes chamada, com mais precisão ainda, psicologia «computacional» — não ouse dizer «calculatória»). É a este par, e ao projecto que o anima, que é consagrado o essencial desta comunicação.

Esta escolha não implica nenhum juízo quanto ao contributo que outras disciplinas trazem e trarão ao estudo dos fenómenos cognitivos. Acontece simplesmente que o projecto de uma ciência cognitiva tal como é actualmente discutido está historicamente ligado à inteligência artificial e à psicologia cognitiva! A possibilidade de uma ciência cognitiva de um género completamente diferente, que se situaria por exemplo inteiramente no quadro da biologia, não é evidentemente de excluir. Acontece que hoje o observador não tem outro ponto de partida possível, parece-me, que não seja o programa de investigação defendido pela inteligência artificial e a psicologia cognitiva.

Começarei, portanto, por apresentar a sua doutrina, que designarei, por comodidade, pelo nome de *cognitivismo*, com o risco de parecer atribuir-lhe uma unidade e uma estabilidade que ela não tem, pelo menos ainda. Tentarei focar a sua força, assim como as suas dificuldades internas. Em seguida apresentarei

* Daniel Andler é maître de conférences na Universidade Paris-VII, associado ao Centro de Investigação de Epistemologia e Autonomia da Escola Politécnica.

In Jean Hamburger, *A filosofia da ciência hoje*,

Lisboa: Feepub, 1988, pp. 113-139

algumas das objecções que ela levanta, tal como certas aproximações concorrentes. Por fim, indicarei as linhas de força segundo as quais o cognitivismo, sob o efeito da crítica, tanto interna como externa, e da concorrência, tenta actualmente reorganizar o seu programa de investigação.

A ideia de base

Se, em certo sentido, podemos fazer remontar as origens do cognitivismo ao nascimento da filosofia ocidental, ou pelo menos a Platão, é a Hobbes que se atribui geralmente a primeira formulação, digamos mesmo a primeira fórmula do cognitivismo:

Raciocinar não é nada mais que calcular (*Leviatã*, 1651).

Mas o que é realmente calcular? A fórmula mantém uma parte de mistério. Para centrar verdadeiramente a ideia cognitivista é preciso ainda:

- Como Descartes, conceber os pensamentos como representações simbólicas de partes ou aspectos da realidade;
- Como Leibniz, imaginar uma «característica universal», quer dizer, um sistema de cálculo simbólico geral que permita operar, por intermédio de símbolos, sobre o conjunto dos objectos abstractos e concretos e obter a totalidade das proposições verdadeiras como combinações convenientes de símbolos;
- Como La Mettrie, exigir que os cálculos simbólicos efectuados pelo espírito humano resultem de interações mecânicas.

O meu propósito não é o de expor a evolução histórica precisa da ideia cognitivista, mas lembrar somente que ela se enraíza na filosofia moderna. Vamos pois directamente à forma que essa ideia tomou durante os anos 50. Ei-la expressa em termos particularmente firmes: um sistema simbólico material possui os meios necessários e suficientes da acção inteligente geral².

Esta segunda fórmula — cuja estranheza não se deve somente à tradução, que é feita palavra a palavra — é enunciada em 1975 por Herbert Simon, um dos fundadores da inteligência artificial e da psicologia cognitiva³. Iremos tentar compreendê-la.

As origens

Voltemos um instante atrás. O cognitivismo contemporâneo nasce na confluência de três grandes ideias que atingem a maturidade no espaço de uma dúzia de anos, a partir de meados dos anos 30.

A primeira é a da *linguagem* ou *sistema formal*. É a noção central da lógica matemática, criada por Frege, Russel e Whitehead, Hilbert, e cuja configura-

ção actual resulta dos trabalhos realizados por Gödel, Tarski, Church e outros nos anos 30.

Um sistema formal é uma espécie de jogo de construção compreendendo diferentes categorias de peças. Por um lado, essas peças arrumam-se segundo regras que respeitam a sua *forma*. Por outro lado, cada peça é um *símbolo*, quer dizer, que tem um *sentido*. E o jogo é de tal modo concebido que, quando as peças são arrumadas em virtude da sua *forma*, constituem um símbolo composto portador de um sentido que resulta da combinação apropriada dos sentidos dos seus componentes. Esta formulação é vaga, mas podemos dar-lhe toda a precisão desejável. Retenhamos que um sistema formal funciona em dois níveis, o da *forma* (nível *sintáctico*) e o do *sentido* (nível *semântico*), e que o seu interesse resulta do paralelismo que se estabelece entre esses dois níveis. Retomando a imagem do jogo, podemos considerar, muito esquematicamente, um *puzzle* como um pequeno sistema formal: o fragmento de imagem impresso sobre cada peça constitui o sentido, e a forma é, literalmente, a forma geométrica. Podemos imaginar um amador que, para complicar as coisas, voltasse as peças e as arrumasse sem ver as imagens que elas têm; ele teria a certeza de, virando em bloco a arrumação obtida, ver aparecer a imagem coerente resultante da combinação de fragmentos de imagem.

A segunda grande ideia é a da *máquina de Turing*⁴. Falta-nos espaço para descrever, mesmo sumariamente, esse objecto, que de resto é relativamente familiar. Portanto, direi simplesmente, alterando a ordem cronológica, que se trata de um computador numérico abstracto. Para o cognitivista, as máquinas de Turing apresentam-se sob quatro aspectos essenciais:

— São *máquinas*, quer dizer, objectos (potencialmente) materiais e obedecendo enquanto tal somente às leis da física;

— São *autómatos*: uma vez postos em marcha, realizam o seu percurso sem intervenção externa;

— A sua *função* é a de operar sobre os símbolos de um cálculo, ou seja, mais uma vez, mediante codificação, *calcular* funções em variáveis e em valores inteiros;

— Entre as máquinas de Turing, algumas, chamadas *universais*, possuem a notável propriedade de poderem calcular não importa que função que o homem, ou uma máquina qualquer, seja capaz de calcular.

Finalmente, a terceira grande ideia é a da *informação*. A teoria matemática que Shannon expõe em 1948 (apoiando-se em ideias que circulavam desde há uma boa vintena de anos) tem por objecto caracterizar a transmissão ou comunicação através de canais imperfeitos de uma grandeza definida de maneira estatística que, pelo menos à primeira vista, não correspondem senão muito parcialmente à noção intuitiva de informação. Uma das escolhas fundamentais do cognitivismo é precisamente considerar como primeira esta noção empobrecida de informação

(sendo a noção comum quer votada a desaparecer, quer reconstituída pela teoria chegada a um grau suficiente de desenvolvimento)⁵.

Aquilo de que são portadoras as peças ou símbolos dos sistemas formais considerados pelo cognitivismo não é, pelo menos em primeira instância, um sentido no sentido habitual — portanto, em particular tal como o entende um sujeito —, mas de uma informação compreendida objectivamente, quer dizer, apreendida como um objecto ou um estado de facto da natureza.

Desenvolvimento da ideia de base

Eis-nos chegados à altura de explicitar a fórmula de Simon:

- a) Pensar, é tratar da informação — a inteligência é uma propriedade dos sistemas de tratamento da informação;
- b) Tratar da informação é calcular, quer dizer, manipular símbolos — os sistemas inteligentes são simbólicos;
- c) Os símbolos manipulados são objectos materiais — os sistemas inteligentes são máquinas que reúnem símbolos passando por uma série de estados que se encadeiam causalmente (em particular automaticamente), segundo regras que dependem somente da arquitectura do sistema, por um lado, e da forma física dos símbolos, por outro lado;

d) Os símbolos são portadores de um sentido, que é uma informação: a reunião dos símbolos segundo a sua forma leva a uma reunião coerente das informações e é por meio desta coerência informacional ou semântica que se manifesta a inteligência do sistema⁶.

Insistamos ainda sobre dois pontos. *Primo*, Simon fala de condição necessária e suficiente; o seu enunciado exprime duas teses ao mesmo tempo: a primeira é que o espírito humano é um sistema simbólico material, a segunda que todo o sistema simbólico material (subentendido: suficientemente complexo) é capaz de inteligência. *Secundo*, um tal sistema é capaz de inteligência *geral*, quer dizer, não somente de uma ou certas tarefas particulares que requerem inteligência, mas, pelo menos potencialmente, de qualquer tarefa desse tipo.

Assim explicado, o enunciado de Simon é perfeitamente límpido? É plausível? Nada menos certo. Mas os cognitivistas consideram-no como o anúncio de um programa de investigação científica, experimental e teórico, não (em princípio) como uma verdade de Igreja. É pois esse programa que se trata de examinar.

A inteligência artificial

A inteligência artificial (ou IA — a sigla está consagrada tal como em inglês a de AI) ocupa no seio das ciências cognitivas um lugar muito particular.

A IA existe, a questão nem se levanta. É um empreendimento técnico-científico já considerável e cujo desenvolvimento, pelo menos quantitativo, se acelera desde há alguns anos a um ritmo vertiginoso. Alguns têm como adquirido que a IA está em vias de provocar a maior reviravolta que a espécie humana jamais conheceu⁷. Posta de lado toda a especulação prospectiva, foi a IA que deu ao cognitivismo o seu impulso inicial: a própria expressão inteligência artificial, imposta por John McCarthy em 1956 (a colegas um pouco reticentes, como Simon⁸), é muito anterior às de psicologia cognitiva, ciência cognitiva ou cognitivismo.

Contudo, a entidade designada pelo termo está mal definida. Uma vez a IA identifica-se com o empreendimento cognitivista no seu conjunto: ela é ao mesmo tempo doutrina e teoria do homem, englobando potencialmente todas as disciplinas tradicionais ou recentemente relacionadas com os fenómenos cognitivos e constituindo, a prazo, uma ciência unificada comparável à física. Outras vezes a IA concebe-se, mais estritamente, como a ciência dos sistemas de tratamento da informação, humanos e artificiais. Outras vezes, enfim, a IA apresenta-se como uma tecnologia, como uma arte ou uma ciência do engenheiro: ela é então engenharia, ou engenharia cognitiva (*knowledge engineering*, em inglês).

Seja qual for, segundo a escola e o período, a ambição que anima a IA, os seus objectivos imediatos correspondem sempre à descrição mínima seguinte: uma tarefa precisa, cuja execução pelo homem requer, intuitivamente, inteligência, sendo escolhida, trata-se de descrever um programa que permita a um computador executar a mesma tarefa com resultados, ou uma taxa de realização, comparáveis aos obtidos pelo homem.

Há que distinguir dois casos. Para determinados problemas, existe um algoritmo, quer dizer, um método de cálculo que conduz infalivelmente ao resultado, quaisquer que sejam os dados, e esse algoritmo é suficientemente simples para ser executado num tempo razoável nas máquinas de que se dispõe⁹. É *grosso modo* desse género de problemas que se ocupa a informática clássica.

É tentador excluir da discussão esta categoria de problemas: independentemente do facto de o seu tratamento por máquina mobilizar por vezes vastos recursos intelectuais (humanos), eles não parecem resultar da inteligência propriamente dita. Quem iria imaginar que uma «calculadora» é inteligente? Os informáticos gostam de sublinhar a «estupidez» dos computadores, pensando por vezes desmistificá-los assim aos olhos do homem médio, quando muito frequentemente confirmam o contrário nas suas opiniões.

A situação não é, no entanto, tão simples como se poderia pensar. Como Russel chamou à atenção, Aristóteles via nas nossas capacidades aritméticas o

signal da nossa racionalidade; e agora que fomos, nesse plano, alcançados (e ultrapassados) pelas máquinas, devemos procurar outros sinais da nossa superioridade¹⁰. Brincadeira? Talvez. Mas, sofisma seguramente, pensarão alguns. Com efeito, *primo*, a máquina não compreende verdadeiramente o que faz: ela não sabe o que é um número ou uma lei aritmética; é abusivo dizer que ela obtém um resultado — é o observador que interpreta convenientemente o seu estado final. *Secundo*, se a máquina atinge esse estado final, ela deve-o não a ela mesma mas ao programador, e o homem não é mais «derrotado» pela máquina que o pianista pelo disco. *Tertio*, que uma máquina consiga chegar ao mesmo resultado que o homem não prova que tenhamos de os considerar como intelectualmente equivalentes; pode acontecer que um seja inteligente e o outro não; permanecendo no domínio da aritmética, entre os calculadores prodígio, alguns são débeis comprovados, outros excelentes matemáticos que gastam nos seus cálculos mentais tesouros de inteligência¹¹.

Estas objecções não deixam de ser importantes, e é significativo que possam ser levantadas a propósito de objectos tão rudimentares e (exceptuando a electrónica) tão antigos como as calculadoras certas questões que alimentam o debate «em verdadeira grandeza». Estas primeiras objecções fornecem-nos de qualquer modo ocasião para contribuir com alguns elementos de clarificação.

Começemos pela mais simples das objecções, a última. É verdade que uma mesma função pode ser calculada e uma mesma tarefa cognitiva realizada, por métodos completamente diferentes. Podem, portanto, ser procurados dois tipos de resultados: ou bem que nos contentamos com programas que conduzem aos mesmos resultados que aqueles que o homem obtém, ou além disso exigimos que os cálculos sejam no essencial os mesmos. Simples igualdade dos grafos no primeiro caso; isomorfismo dos procedimentos no segundo. Falaremos, respectivamente, de equivalência *fraca* ou *forte*¹², entre o programa da máquina e, se é que isso tem algum sentido, o «programa» humano. Se importa distinguir essas duas formas de equivalência, não temos por isso fundamentos para desprezar qualquer resultado de equivalência fraca.

A segunda objecção (segundo a qual a pretensa inteligência da máquina não é mais que uma cópia da do programador) pede um primeiro reparo, aparentado com o precedente: é preciso reconhecer que entre a cópia e o original existe uma dissimetria radical; donde os dois tipos de objectos concebíveis, fabricar uma cópia, fabricar um original, donde pode resultar que um seja acessível e o outro não. Mas a fabricação de uma cópia, longe de ser negligenciável, pode representar um resultado importante. No plano prático, em primeiro lugar, mesmo uma «simples» cópia pode ter a sua utilidade; além disso, uma cópia não é necessariamente «simples» — ela pode, por exemplo, amplificar o original segundo determinadas dimensões, incorporar características das quais ele está desprovido, entrar em combinações para as quais por diversas razões ele pode não se prestar.

No plano teórico, por outro lado, a possibilidade de copiar um processo intelectual dado não é de modo algum indiferente, pois ele revela frequentemente um aspecto suficientemente encoberto para que um Pascal, por exemplo, não desdenhe estudá-lo; e que, além disso, traz um elemento de resposta a um problema do maior interesse, o de caracterizar, entre as diligências de que o homem parece capaz, aquelas que são *copiáveis*, reproduzíveis por uma máquina.

Finalmente, a primeira objecção (segundo a qual não se poderia literalmente sustentar que a máquina «compreende o que faz») é sem dúvida fundamental, e voltaremos a ela. Mas é preciso desde já notar que, até certo ponto, tudo se passa *como se* a máquina compreendesse, quer dizer, como se ela fosse capaz de inferir um sentido (um resultado numérico) a partir do sentido (do valor) dos dados. Não é indiferente que sejamos pelo menos tentados, para explicar o «comportamento» (o funcionamento) de uma calculadora, a atribuir-lhe um modo, um fantasma de pensamento. Um objecto que se presta a um tal discurso não é por certo necessariamente uma inteligência; também não é um objecto qualquer.

A escala muito reduzida em que nos colocámos por um instante, estas considerações podem parecer fúteis; elas assumem talvez um outro peso desde que pensemos nos engenhos complexos que a inteligência artificial já produz ou espera produzir mais cedo ou mais tarde.

Voltemos agora ao segundo género de problemas, aqueles para os quais não existe algoritmo, pelo menos de complexidade razoável. É tipicamente assim que se apresenta um problema de IA. Neste caso, procura-se definir uma estratégia não necessariamente com êxito. Incorpora-se assim ao programa um conjunto de regras que lhe permitem efectuar em diferentes estados, em função da situação do momento, certas escolhas que favorecem, mas sem a garantir, a descoberta de uma solução. Um programa deste género diz-se *heurístico*. A aposta da IA pode então ser assim formulada: qualquer tarefa cognitiva pode ser realizada por um computador numérico programado heurísticamente.

Esta aposta foi mantida? Naturalmente que não. Sê-lo-á um dia? Pergunta bastante mais difícil, à qual foram dadas todas as respostas concebíveis. Investidores e observadores estão no entanto de acordo em admitir que os obstáculos são consideráveis, os resultados frágeis, os progressos lentos e, em certas direcções, nulos. Mas devemos ficar espantados? Que conclusões tirar daqui? Para uns, essas dificuldades são o sinal de um programa de investigação ainda mais apaixonante do que tinham previsto. Para outros, que esse programa deve ser abandonado, no essencial, ou mesmo que nunca devia ter sido levado a sério. Para outros, enfim, que ele deve ser completamente repensado no quadro de uma metodologia rigorosa e à luz de alguns resultados, positivos mas também negativos, que são realmente significativos. Para tentar julgar, é preciso olhar mais de perto: quais são então as dificuldades com as quais se confronta a IA?

Uma abordagem, muito sumária para informar, mesmo que parcialmente, um verdadeiro juízo, permitir-nos-á, contudo, compreender alguns lances.

Dificuldades pragmáticas do projecto cognitivista

Nos primeiros tempos a IA pôs o acento sobre a modelização de capacidades gerais de resolução de problemas. Três hipóteses orientavam esses trabalhos:

a) O essencial da actividade cognitiva consiste em resolver problemas;
b) Existem no homem mecanismos gerais de resolução de problemas, que se aplicam a todos os domínios;

c) Exibir-se-ão esses mecanismos observando sujeitos humanos ocupados em resolver diversos problemas e inspirar-nos-emos nessas observações para construir programas heurísticos, permitindo o computador ao mesmo tempo verificar as conjecturas incorporadas nesses programas e generalizar eventuais soluções.
Esses primeiros trabalhos conduziram indiscutivelmente a alguns resultados. Mas as soluções propostas só incidiam sobre um tipo muito estritamente limitado de problemas excessivamente simples, e as soluções propostas não eram generalizáveis nem no sentido de uma grande complexidade, nem no de uma maior variedade das tarefas. Além disso, essas soluções permaneciam vazias de conteúdo teórico, na medida em que, contentando-se em mimar certos comportamentos sem explicar previamente o problema científico a resolver e nomeadamente os parâmetros, era difícil dizer o que realmente esses resultados provavam.

A IA procurou então redefinir o seu programa de investigação quer revendo por baixo os seus objectivos quer enriquecendo a sua panóplia de conceitos fundamentais.

Para os objectivos foram sucessivamente considerados três tipos de limitação:

— Renunciar à simulação dos procedimentos humanos para se voltarem para a construção de aparelhos capazes de resultados comparáveis, até mesmo superiores (e não transformações internas idênticas): já não se visa a equivalência forte, mas a equivalência fraca; renuncia-se à inteligência sintética para procurar os princípios de uma inteligência propriamente artificial;

— Renunciar a uma inteligência versátil, e limitar-se a programas estritamente especializados numa tarefa dada;

— Renunciar a uma inteligência geral e limitar-se a programas operando somente num contexto estritamente (e previamente) limitado.

Entre os seus conceitos fundamentais a IA admitiu sucessivamente:

— Uma semântica dos objectos manipulados — a informação tratada pelo sistema deve sê-lo em função do «sentido» dos símbolos, quer dizer, de um conjunto de conhecimentos específicos que determinam a função dos objectos simbolizados;

— Uma noção de mundo, de contexto, no qual a inteligência, quer seja humana ou artificial, se exerce, e que se caracteriza por uma diversidade qualitativa de dados e tarefas; espaço físico, objectos materiais móveis, percepção, comunicação, deliberação, justificação, planificação, acção, verificação, etc.;

— Um *corpus* estruturado de conhecimentos gerais — toda a acção inteligente num contexto que não é artificialmente limitado apela a múltiplos conhecimentos de ordens muito diversas, pertencentes em particular ao senso comum ou ainda às práticas do falar e, mais alargadamente, comunicativas.

Naturalmente, à medida que a IA enriquecia a cognição de fenómenos sempre mais complexos, tinha necessidade, por realismo, de reduzir a escala do problema a resolver. A noção de micromundo, desenvolvida no início dos anos 70, ilustra bem este duplo encaminhamento: tratava-se de universos artificiais reduzidos segundo todas as suas dimensões: variáveis, situações e acções em número limitado (e predeterminado), contextos fechados, linguagem «natural» mas restringida a um muito pequeno número de funções sintácticas, semânticas e pragmáticas, reflexão e justificação de tipo uniforme e predeterminado, etc.

A questão era então de saber em que é que as soluções elaboradas para esses micromundos (por exemplo, o funcionamento inteligente num micromundo) eram, senão gerais, pelo menos inteligíveis e generativas, prestando-se portanto, por refinação e composição, à elaboração gradual de soluções em tamanho natural de problemas cada vez mais próximos dos verdadeiros problemas.

A esta questão foi preciso dentro em pouco responder de maneira bastante negativa. A IA quis então escapar à armadilha do caso particular levantando o problema da representação dos conhecimentos. Os homens são manifestamente capazes, em cada situação, ou quase, de mobilizar quase instantaneamente, no seio de um *stock* gigantesco de conhecimentos, quer dizer de factos, regras, regras de aplicação das regras, etc., aquelas de que necessitam: eles utilizam somente o que é «pertinente», sem serem obrigados a passar em revista o que não o é. Como é preciso organizar um vasto repertório de conhecimentos de maneira a poder extrair deles a qualquer momento o que é necessário para o tratamento de uma situação emergente de um espaço infinito de possibilidades? Os esquemas propostos pela IA — «guiões», «cenários», «quadros», «MOPs» (*memory organization packets*), etc. — são outras tantas tentativas engenhosas para conciliar flexibilidade e rapidez oferecendo uma escolha variada, mas no entanto limitada, de possibilidades, adaptada à situação.

Pressentimos talvez os limites do procedimento: pode-se predeterminar o mundo humano num número finito, por muito grande que seja, de situações-tipo? É a pertinência uma relação suficientemente estável para se deixar inscrever em filigrana numa estrutura dada, mesmo que ela tenha geometria variável? Limito-me a levantar o problema e indicarei somente que a insuficiência das soluções

propostas é atestada por um enigma que a IA reconhece não ter penetrado e que ela designa pela expressão *problema do quadro (frame problem)*. Vejamos do que se trata: um robô inteligente serve para executar certas tarefas num ambiente complexo; a execução dessas tarefas conduz a uma modificação desse ambiente; para poder continuar a agir inteligentemente, o robô deve estar à altura de registar essas modificações ou, como se diz, de actualizar a sua base de dados. Onde reside o problema? No facto de que, na imensa quantidade de aspectos que poderiam *a priori* ser modificados, muito poucos o são efectivamente e muito menos o são ainda de maneira sensível. Como evitar que o robô que serra o ramo sobre o qual está sentado não esteja ainda a calcular o peso da serradura produzida e a decidir desprezá-la no momento em que o ramo se parte? Ninguém tem, de momento, resposta satisfatória para esta pergunta.

Começamos a perguntar-nos, em IA, se a construção de sistemas «verdadeiramente» inteligentes, quer dizer, abertos e operando em contextos abertos, não exigirá conceitos radicalmente novos.

Enquanto prosseguem estas reflexões teóricas, mesmo francamente especulativas, a engenharia do conhecimento (*knowledge engineering*) desenvolve-se consideravelmente, ajudada por uma visão relativamente clara dos seus objectivos e dos seus métodos. Os seus objectivos são os de desenvolver sistemas especializados operando o melhor possível em domínios de tipo «micromundo» apresentando um real interesse técnico económico ou social. Os seus métodos são os da resolução de problemas (nomeadamente a dedução automática) e da representação dos conhecimentos.

O mais notável resultado dessa iniciativa é a ideia de sistema pericial. Aí reside um triunfo técnico da IA, mas também, curiosamente, quase uma recusa teórica, ou pelo menos a renúncia — provisória? — a uma boa parte das suas ambições. Com efeito, não somente o domínio de aplicação de um sistema pericial é um micromundo particularmente limitado e completamente descontextualizado, mas além disso os conhecimentos que utiliza, longe de serem inferidos a partir de alguns princípios inteligíveis e algumas séries de dados, são muito simplesmente armazenados na desordem sob a forma de centenas ou milhares de regras obtidas, no decurso de um processo de parto do qual não se sabe o que quer que seja do alcance geral, junto de especialistas humanos, quer dizer de especialistas do domínio considerado. O que torna a coisa tecnicamente possível, é que a representação dos conhecimentos é simplificada pela eliminação do contexto em sentido lato.

O que não é naturalmente um magro resultado. Isto não impede que o sonho de um computador que, posto em contacto com um ambiente real, adquiriria por si mesmo os conhecimentos para o seu próprio progresso continue a ser um sonho. Eis, portanto, a IA levada ao cimo da notoriedade pelo efeito do seu mais modesto sucesso.

Contudo esse sucesso não deixa de ter alcance teórico; uma intuição da IA moderada aparece com efeito fortalecida: em certos casos, muitos conhecimentos, donde quer que provenham, e sem qualquer razão, bastam para reproduzir pelo menos aproximativamente uma função que exige no homem inteligência¹². É mesmo o tipo de resultado do qual eu atrás sugeria (a propósito das calculadoras) que está longe de ser desprezável.

A psicologia cognitiva e os problemas teóricos do cognitivismo

Contrariamente à IA, disciplina surgida quase *ex nihilo*, a psicologia cognitiva teve de justificar a sua instalação no seio de uma disciplina constituída. Teve de escorar a sua pretensão de se substituir a um paradigma que se esvaziava, o behaviorismo, através de uma análise dos seus próprios fundamentos. Foi por isso que a discussão epistemológica mais cerrada do cognitivismo se desenrolou na proximidade da psicologia cognitiva.

Há outras diferenças relativamente à IA. O computador continua a ser, em psicologia cognitiva, uma referência conceptual e uma ferramenta de investigação e verificação, mas o lançamento de uma teoria sob a forma de programa (de computador) não é um imperativo. Pelo contrário, sendo o objecto de estudo a cognição humana, não se trata, a não ser talvez a nível de etapa intermediária, de se contentar com uma equivalência fraca entre teoria e processo humano — somente a equivalência forte terá um alcance científico. Por outro lado, não obstante a minha terminologia demasiado esquemática, a psicologia cognitiva não está inteiramente ligada pelos postulados do cognitivismo: guarda programas de investigação que se diferenciam dela em maior ou menor grau. Mas não insistirei sobre essas distinções, e dispensar-me-ei mesmo (por razões de espaço) de uma descrição, mesmo que sumária, do desenvolvimento e dos principais eixos de investigação da disciplina¹⁴. Em contrapartida tentarei expor como a psicologia cognitiva procurou explicitar os postulados do cognitivismo, ou ainda, o que vem a dar no mesmo, precisar com que filosofia de espírito ela parece compatível.

O funcionalismo

O ponto de partida da psicologia cognitiva é um triplo postulado:

- a) Existem, na realidade, estados mentais e esses estados mentais entram nos processos mentais;
- b) Os estados mentais têm uma existência material em tanto que estados físicos, por exemplo neuronais, e os processos mentais são processos físicos;
- c) A explicação psicológica desenvolve-se a um nível que lhe é próprio, e que

não é principalmente redutível ao da explicação física (mesmo *via* biologia e química): uma lei ou uma generalização psicológica incide sobre as relações que mantêm os estados mentais definidos enquanto tal.

É esta concepção da explicação psicológica que se designa desde há uma quinzena de anos pelo termo *funcionalismo*.

Para compreender plenamente o alcance desta doutrina, pode ser útil precisar, como o faz Jerry Fodor¹⁵, ao que ela se opõe:

— Ao atribuir um papel explicativo aos estados mentais, opõe-se ao behaviorismo, que pura e simplesmente os elimina em proveito de pares estímulo-reacção;

— Ao considerar um estado mental como um estado físico, opõe-se ao dualismo;

— Ao afirmar a existência de um nível de explicação propriamente psicológico (ou, o que vai a dar no mesmo, ao defender que os estados mentais entram *enquanto tal* em certas relações), opõe-se ao monismo reducionista, quer dizer, à redução do psíquico ao físico.

Poder-se-ia pensar que os dois últimos pontos são inconciliáveis. De modo nenhum. Deve-se com efeito distinguir dois tipos de monismo, o «fraco» e o «forte». A tese monista fraca é a da *identidade ocasional (token identity)*: cada estado mental particular — tal estado de tal indivíduo em tal momento — é idêntico a (ou realizado por) um estado físico, neuronal por exemplo. A tese monista forte postula, para toda a classe de estados mentais que intervêm nas explicações da psicologia, a existência de um predicado da física (ou da neurofisiologia) que lhe é co-extensivo. Há portanto aqui *identidade categorial (type identity)* entre os estados mentais e certos estados físicos.

É pelo menos logicamente possível, como se vê, adoptar a tese da identidade ocasional rejeitando ao mesmo tempo a da identidade categorial. Mas temos razão para o fazer? Alguns filósofos, entre os quais Fodor, pensam que sim. Eis, à laia de argumento, dois exemplos tendentes a apoiar esta posição.

Tomemos em primeiro lugar o conceito administrativo de circunscricção. É certo que qualquer circunscricção é idêntica (aceitemos) a um domínio fisicamente determinado. Mas podemos imaginar que existe uma categoria fixa (por exemplo geográfica) natural, quer dizer, que intervém de maneira natural nas generalizações da física (ou da geografia física), correspondendo à de circunscricção eleitoral? Não é absolutamente impossível, mas não é muito plausível: seria necessário admitir que uma tal categoria entra enquanto tal num esquema de explicação que liga, por exemplo, tal tipo de mudança de regime político a tal tipo de modificação do corte eleitoral.

Consideremos agora a noção de saca-rolhas. Um saca-rolhas é seguramente um objecto físico. Mas a noção de saca-rolhas intervém em generalizações que incidem sobre os piqueniques, o alcoolismo, o comércio ou os ritos, nos quais não

intervém o tipo de saca-rolhas considerado — de simples ou dupla hélice, de lâminas, de gás comprimido, de rosca ou de alavanca, etc. Uma categoria física que poderia tomar o lugar da do saca-rolhas seria, portanto, descrita por uma pouco verosímil disjunção de especificações de aparelhos muito diversos, que, para cúmulo, se alargaria em cada entrada no domínio dos saca-rolhas! Somos então tentados a distinguir uma *função* sociogastronómica de saca-rolhas das suas múltiplas materializações: é o que faz o funcionalismo na psicologia, substituindo as circunscricções e os saca-rolhas pelos estados mentais.

O cognitivismo é, portanto, funcionalista; mas há mais: o seu funcionalismo é *turingiano*. Isto significa que o sistema funcional no qual os estados mentais têm lugar é semelhante a uma máquina de Turing. Por outras palavras, as funções que eles preenchem devem ser decomponíveis em funções de simplicidade crescente, sendo as funções mais simples isomorfas às operações elementares de uma máquina de Turing. Esta obrigatoriedade suplementar é essencial para garantir o poder explicativo da teoria: ela proibe ao funcionalista fazer figurar na sua teoria componentes tais como um «resolutor universal de problemas», como um «colector de informações pertinentes», como um «sintetizador de obrigações» ou ainda um «gerador de hipóteses interessantes», a menos que tenha decomposto os referidos componentes em componentes elementares. O funcionalismo turingiano está, além disso, protegido contra qualquer recaída na tese da identidade categorial: com efeito, ele procura uma descrição dos *procedimentos* mentais que faz intervir de maneira canónica a sua função psicológica e não as suas propriedades físicas.

As representações mentais

Que são os estados mentais que a explicação psicológica cognitivista põe em jogo? O funcionalismo não se pronuncia sobre esta questão e parece compatível com diversas respostas. O cognitivismo, no entanto, faz aqui apelo a uma hipótese nova, pelo menos parcialmente independente do esquema funcionalista: um estado mental é, ou caracteriza-se por, uma *representação*. Mas é claro que uma representação continua a ser inerte enquanto não for, de uma maneira ou de outra, tomada em consideração pelo sujeito que a tem ou a produz. Deve-se, portanto, de maneira mais precisa, ver um estado mental¹⁶ como uma *relação* entre o sistema (ou o sujeito) e a sua ou as suas representações. O cognitivismo parece assim dever ceder o lugar à *intencionalidade* — a que voltaremos um pouco mais adiante.

Que se deve entender por representação? Questão decisiva, já que todos os programas de investigação de inspiração cognitivista, dos mais abstractos aos mais concretos, apelam mais ou menos explicitamente a essa noção.

Reflexões
O cognitivista moderno não pode pensar as representações como *imagens da realidade*, sob pena de ter de postular um homúnculo que as contemplaria e as analisaria a fim de inferir delas as informações de que o sistema (o sujeito) precisa: o homúnculo, por seu turno, formaria as representações internas, donde a regressão infinita. Por outro lado, conceber as representações como sensações, ou combinações de sensações, rebaixaria o cognitivista a uma solução radicalmente fiscalista, quer dizer, no estado presente do conhecimento, a um mistério que é pouco plausível que se dissipe rapidamente, ou mesmo nunca.

Uma representação mental, para poder preencher o papel explicativo geral que o cognitivista lhe quer atribuir, deve ser entendida como uma *descrição simbólica*. Deste modo, a sequência das representações mentais torna-se em algo que se assemelha a um conjunto de enunciados de uma linguagem formal: em virtude da sua forma, que é material, esses enunciados entram em transformações cuja série constitui o conteúdo de um processo mental; em virtude do sentido de que são portadores, eles fornecem ao sistema informações sobre o mundo (do qual faz, aliás, parte).

Sentido e intencionalidade

Reaparece, portanto, aqui a questão do sentido, evocado a propósito das máquinas de calcular, para ser de imediato posto entre parêntesis. Não podemos continuar a ignorá-lo.

Como faz a representação sentido para o sistema? Um estado mental, como vimos, é uma representação para o sistema: é uma «atitude proposicional»¹⁷, quer dizer, uma posição (crença, dúvida, negação, interrogação, esperança, finalidade...) do sistema frente às combinações simbólicas que constituem as suas representações, entendidas naturalmente no seu *sentido*. O comportamento do sistema não é verdadeiramente *inteligível*, quer dizer, não contém essa marca da inteligência que é a *racionalidade*, a não ser na condição de que a atribuição de um sentido às formas simbólicas seja o facto não (somente) do observador (ou, no caso de uma máquina, do seu fabricante), mas originalmente o do próprio sistema. Como explicar, sem regressão, esta atribuição do sentido por um organismo material que as suas representações não podem causalmente afectar a não ser em virtude da sua forma material?

A oposição filosófica ao cognitivismo (Dreyfus, John Searle) considera que nenhuma resposta é concebível nesse quadro e conclui daí a sua incoerência de fundo. Entre os partidários ou simpatizantes da doutrina alguns procuram resolver ou dissolver o problema; mas outros (Fodor, Robert Cummins) consideram que ele não é eliminável nem está resolvido. Cummins, um dos mais rigorosos entre os filósofos próximos do cognitivismo, analisa as diferentes tentativas de solução

e rejeita-as uma a seguir à outra. Ele mostra em particular por que motivo a estratégia cognitivista de redução progressiva da inteligência ao seu grau zero é aqui impotente: a inteligência é produzida (a acreditarmos nos cognitivistas), a intencionalidade não. O mistério do sentido originário permanece completo.

No entanto, sugere Cummins, isso não deve impedir o programa cognitivista de desenvolver uma teoria da «pseudocognição»¹⁸, com base numa espécie de «crédito de sentido» consentido pelo observador ao sistema observado e pelo epistemólogo à teoria nascente. Para Cummins, como para Fodor, esta teoria conserva uma hipótese de vir a ser uma teoria da cognição propriamente dita, no dia em que encontrar o meio de pagar a sua dívida; nela reside, segundo eles, a única esperança do cognitivismo. Há nisto uma espécie de filosofia do «como se», mas que pensa não continuar assim: ela demitir-se-á — anuncia — se a ciência se mostrar incapaz de honrar a sua dívida.

Concepções e aproximações concorrentes

Restringi-me até agora ao quadro teórico definido pelo cognitivismo, adoptando o seu ponto de vista e evocando apenas aquelas suas dificuldades que se deixam apreender no seio da própria teoria. É altura de mencionar as críticas formuladas de fora, assim como os programas de investigação dissidentes ou concorrentes.

Críticas globais da aproximação cognitivista

Das inúmeras críticas dirigidas ao cognitivismo podemos isolar, simplificando, quatro grandes estratégias de rejeição.

a) Podemos considerar que o cognitivismo é *conceptualmente incoerente*, nomeadamente:

— porque deixa intocável o enigma do sentido original e continua a ser incapaz de explicar que um sentido possa ter um efeito causal no mundo (Searle);

— porque postula, entre o nível físico (ou neuronal) e o nível fenomenológico, o da experiência quotidiana, um nível informacional ou simbólico do qual nada prova a existência — ou ainda porque a noção de representação que está no âmago da teoria é ininteligível (Ryle, Dreyfus);

— porque a própria noção de informação, aplicada a um ser vivo mergulhado num meio, ou a um ser humano mergulhado num universo humano, não pode ser transposta a partir da física — ela deve ser entendida completamente ao contrário (Atlan, Maturana, Varela, Flores e Winograd, Dreyfus);

— porque a concepção de um universo objectivo feito de elementos pré-definidos combinando-se segundo regras determinadas, concepção fecunda na física, e considerada pela corrente dominante da filosofia ocidental como aplicável ao mundo humano, é completamente ilusória (Dreyfus, Von Foerster, construtivismo de Palo Alto, etc.).

b) Podemos considerar que o cognitivismo é manifestamente *incompatível com a experiência fenomenológica* (eventualmente clarificada por resultados da psicologia experimental, nomeadamente, mas não exclusivamente, de inspiração «gestalista»). Que, por exemplo, a experiência nos mostra que o sentido não é composto segundo regras a partir de átomos de sentido, que ele é global (holista), contextual, emergente. Que as nossas capacidades cognitivas estão indissoluvelmente ligadas à nossa propriedade de ter um corpo (e não somente um cérebro). Que a nossa inteligência é à primeira vista a de um ser vivo situado num meio, portanto atravessado por um feixe variável de intenções, de temores, de finalidades. Que nós somos irremediavelmente seres sociais, mergulhados desde o nascimento num horizonte indefinido de práticas sociais. Que deste modo qualquer actividade pressupõe um contexto, e que um contexto é uma coisa completamente diferente de um conjunto de factos ou de regras que modificam ou completam os factos e as regras relativos a não se sabe que «estado neutro» do «sistema». Que nós só muito raramente nos ocupamos em resolver problemas e que mesmo essa actividade consiste, muito frequentemente, não em produzir inferências, mas em reconhecer uma situação, quer dizer, em identificá-la como semelhante a um protótipo escolhido entre milhares ou dezenas de milhares. Que o modo como nós adquirimos e depois dominamos uma actividade, tal como tocar um instrumento, jogar xadrez, manejar uma ferramenta, conduzir um automóvel, receber convidados, etc., mostra que nós não procedemos, excepto talvez no início, pela aplicação de regras, por muito complexas que sejam, a factos elementares. Ou, enfim, que os humores e as emoções, na medida em que elas afectam globalmente o conjunto das nossas disposições cognitivas, escapam ao esquema cognitivo (Dreyfus, Haugeland, Flores e Winograd, Rosch)¹⁹.

c) Podemos suspeitar que o cognitivismo é *incompatível com a realidade biológica*. Muitos investigadores, biólogos, mas também um informático como Douglas Hofstadter ou um filósofo como Daniel Dennett, não vêem de modo algum surgir, através da estrutura geral de explicação proposta pelo cognitivismo, uma solução biologicamente plausível. Os esquemas cognitivistas são demasiado económicos, demasiado austeros, demasiado hierarquizados, demasiado perfeitados. De modo mais preciso, alguns neurobiólogos duvidam que uma tal estrutura possa ser «implementada», materializada no *wetware*, no tecido cerebral.

Tais argumentos, presentemente, não são decisivos — se o fossem, isso dever-se-ia ao cognitivismo. Mas encorajam alguns investigadores a explorar vias muito divergentes.

d) *Finalmente, podemos considerar que o cognitivismo é logicamente contraditório*, na medida em que, por exemplo, o seu sucesso implicaria a existência, pelo menos virtual, de um ser dotado de um autoconhecimento perfeito. Tentou-se igualmente aplicar os teoremas de incompletude de Gödel à refutação de toda a tese mecanicista que procura identificar um ser humano (mais precisamente um matemático) com uma máquina de Turing. Mas estes argumentos parecem ser contestáveis, quer na sua forma quer no seu alcance supostos²⁰.

Programas concorrentes: críticos e rebeldes

Numa panorâmica excessivamente rápida das aproximações que se desenvolvem desde há alguns anos nas fronteiras do cognitivismo, ou em franca oposição com ele, distinguirei, para simplificar, o crítico, que permanece fiel aos seus principais postulados, e a rebelião, que rejeita alguns.

Os críticos põem em causa o carácter exclusivamente discursivo, proposicional, da informação tratada pelo espírito humano. A velha ideia de imagem mental regressa em força — sendo a hipótese que nós formamos tais imagens e que as exploramos, ou as manipulamos, directamente, sem tradução prévia num idioma simbólico, à medida das nossas necessidades²¹. Assim, para verificar se dois volumes orientados diferentemente no espaço são sobreponíveis, fartamos rodar a imagem mental de um para a colocar paralelamente à do outro.

De um modo geral, tentamos dar consistência à hipótese de uma informação analógica, e já não discreta ou numérica²². Mas restringimo-nos, no criticismo, às noções principais, que são a informação abstracta, as representações mentais, os símbolos formais, o paralelismo forma/sentido, o atomismo semântico e a com-posicionalidade do sentido, a resolução de problemas, etc.

É uma ou outra destas noções, ou todas ao mesmo tempo, que os rebeldes recusam, ao procurarem outros fundamentos para um estudo da cognição. Uns (J. J. Gibson, U. Neisser, B. Shanon) colocam-se imediatamente no campo psicológico, cingindo-se particularmente à concepção cognitivista da representação (deste modo, Gibson e os seus continuadores elaboram uma teoria da percepção directa, Shanon concebe a inteligência como oscilando entre um modo representacionista e um modo a-representacionista, etc.).

Outros situam a sua reflexão no quadro geral de um estudo dos sistemas biológicos, vistos como organismos auto-organizados, assumindo assim uma parte da herança da cibernética, sofredoramente eclipsada, durante uma vintena de anos, pela IA em plena ascensão. Assim, a escola chilena (H. Maturana, F. Varela) lançou as bases de uma teoria das máquinas autopoieticas²³. Mas é sem dúvida o conexionismo que mais prende actualmente as atenções, porque alia à procura

de uma nova conceptualização a dos objectos concretos da psicologia cognitiva e da IA.

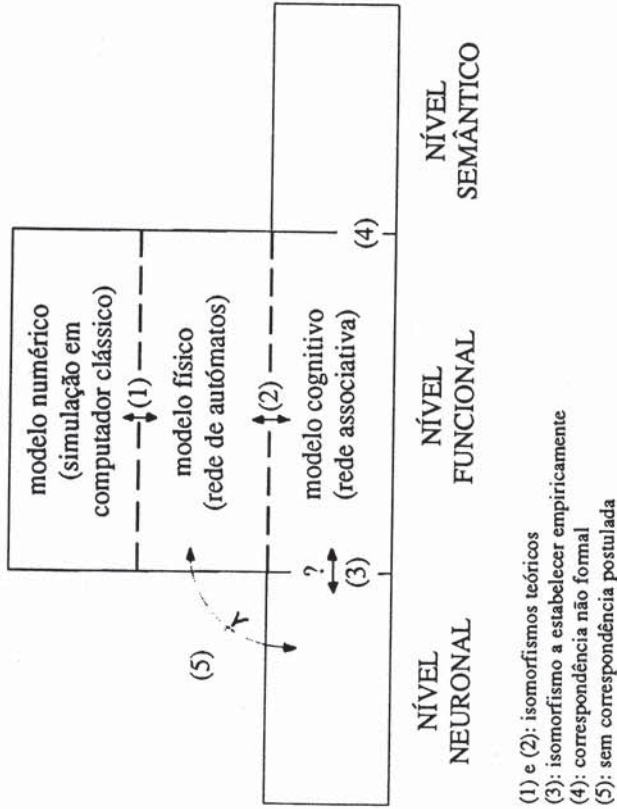
A aproximação conexionista (também chamada por vezes *neocconexionista* porque se inspira directamente nos trabalhos conexionistas compreendidos por McCulloch e Pitts no início dos anos 40 e desenvolvidos por Von Neumann durante a década seguinte) toma como objecto de base *redes* funcionalmente descritas como nós ligados por canais, segundo diversas configurações mais ou menos regulares²⁴. Cada nó pode tomar uma variedade de estados, por exemplo dois (activo/inactivo), e transmitir àqueles ao qual está ligado, segundo o seu estado, uma activação ou uma inibição. Os nós colocados na periferia da rede podem captar activações positivas ou negativas provenientes do ambiente. Uma vez posta em marcha, a rede no seu conjunto passa por uma sequência de estados, e acaba por atingir uma situação oscilante ou uma situação estável. Essa situação estável, se se estabelece, constitui globalmente o valor «calculado» pela rede.

As propriedades funcionais de uma tal rede resultam essencialmente da *força* das ligações ou conexões (daí a denominação do domínio) que se estabelecem entre os nós no decurso de um processo de aprendizagem inicial, depois estabilizam-se e permitem à rede «adulta» assumir o seu papel cognitivo.

Os nós, na interpretação psicológica de uma rede, não estão necessariamente associados a uma unidade ou a um átomo de sentido predeterminado. Em certos modelos (os «associadores de padrões»), a unidade funcional é, além disso, definida de modo puramente global e dinâmico: trata-se de configurações de activação que se desencadeiam mutuamente. É, portanto, da maneira como as redes «tratam» a informação que elas se diferenciam mais nitidamente dos modelos de IA e da psicologia cognitiva ortodoxas; o sentido só aparece ao nível do processo global; não é constituído por átomos de sentido ligados a cada componente, e que entram em composição definida no termo de uma sequência de transformações. Existe realmente uma correspondência entre o nível semântico e o nível funcional, mas essa correspondência não deve ser entendida como um paralelismo formal. Trocou-se, objectar-se-á, um mistério por outro. Talvez: escolher um programa de pesquisa não é um pouco escolher o mistério perante o qual nos sentimos mais atraídos?

Como quer que seja, no conexionismo a ordenação dos níveis apresenta-se da maneira indicada no esquema que se segue.

As redes do conexionismo, como esquemas de descrição psicológica, são chamadas, e compreende-se porquê, redes associativas ou modelos de activação interactiva. Como classe de sistemas materiais, eles são modelizados pelas redes de autómatos finitos de diferentes espécies — os mais correntemente utilizados são os autómatos com limiar, a que McCulloch e Pitts chamavam *neurónios formais*.



- (1) e (2): isomorfismos teóricos
- (3): isomorfismo a estabelecer empiricamente
- (4): correspondência não formal
- (5): sem correspondência postulada

Este último termo, que certamente não foi escolhido por acaso, faz-nos correr o risco de deixar pensar que as redes de autómatos são, pelo menos aproximadamente, modelos neuronais — desempenhando cada neurónio formal o papel de um neurónio verdadeiro e os canais o dos axónios e dos dendritos, etc. Ora é isto que os *neocconexionistas* se abstêm de postular: os seus modelos não são mais que descrições funcionais cuja implementação no tecido nervoso continua indeterminada.

Vemos aparecer no esquema acima um nível suplementar: as redes de autómatos não são estudadas directamente, mas simuladas em computadores clássicos, isto é, de arquitectura sequencial. Um dos principais objectivos do conexionismo é justamente o de definir novas arquitecturas, massivamente paralelas, para os computadores do futuro.

Os modelos conexionistas são objecto de trabalhos cada vez mais numerosos. Eles parecem ser particularmente eficazes no estudo de fenómenos de acesso lexical, de memória, de reconhecimento de formas, de aprendizagem e de controlo motor nas tarefas automáticas tais como a manutenção do equilíbrio ou a dactilografia. Em certos domínios, o conexionismo surge cada vez mais nitidamente como um rival do cognitivismo clássico.

Conceptualmente ele parece, em todo o caso, apresentar vantagens sensíveis. Os seus modelos são, pelo menos intuitivamente, próximos da biologia (fonte de inspiração reconhecida): eles são «neuronóides», à falta de serem neuronais. Uma rede funciona sem controlo executivo, sem programa nem chefe de orquestra. O processo é massivamente paralelo e não essencialmente sequencial. Não tem verdadeira semântica interna, nem transmissão no seio da rede de informações de complexidade crescente. A função cognitiva está disseminada, distribuída. O papel dos componentes não é definido antecipadamente; eles são ambivalentes, simplistas, intermutáveis. O sistema pode funcionar mais ou menos bem e ser danificado sem perder todas as suas qualidades. Pode ser activo sem «rodar», como o faz um programa, funcionar sem resolver um problema: uma rede não procura soluções — as «soluções» emergem ocasionalmente enquanto configurações estáveis. Em suma, àqueles que o cognitivismo clássico deixa cépticos, mas que não renunciavam a uma aproximação empírica da cognição, o neoconexionismo parece ter bastante para dar.

Para uma estruturação do domínio

Factorização e modularidade

Abalado pelas críticas, por alguns falhanços, pelo desafio lançado pelas pesquisas concorrentes, o cognitivismo, sob o impulso de investigadores tais como Fodor, Zenon Pylyshyn, David Marr e outros, procura redefinir o domínio e os métodos que lhe são próprios.

O princípio organizador ao qual o cognitivismo hoje apela consiste em decompor o objecto de estudo em subobjectos *naturais*, o que é, como se sabe, mais fácil de dizer que de fazer: a história da IA dá disso muitas provas. Actualmente são sugeridos dois modos de decomposição.

Podemos, em primeiro lugar, levantar a hipótese de uma *factorização* dos fenómenos cognitivos²⁵. Eles seriam submetidos a duas ordens de determinação: uma determinação enquanto processo cognitivo propriamente dito (no sentido que procura precisar o cognitivismo clássico), e uma determinação pelo tipo de implementação funcional efectivamente realizada no cérebro. A imagem do computador aclarará a ideia: o processo cognitivo é um algoritmo abstracto; o fenómeno cognitivo total resulta simultaneamente do algoritmo e da maneira abstracta como esse algoritmo é executado, a qual depende da máquina virtual utilizada (ou, o que vai a dar ao mesmo, da linguagem de programação na qual o algoritmo é escrito).

O ganho é o seguinte: podemos salvar a ideia-força do cognitivismo continuando a admitir a existência de efeitos cognitivos que não relevam do esquema

cognitivista: são aqueles que resultam de factores ligados à *arquitectura funcional* do sistema. Isolar esses factores não é coisa fácil. Pylyshyn pensa poder caracterizá-los pela sua «impenetrabilidade cognitiva» — visto que eles são insensíveis às condições propriamente cognitivas nas quais se exercem: crenças e intenções do sujeito, contexto. É precisamente a este tipo de efeitos que se aplicariam os modelos neoconexionistas, assim como a teoria dos processos inferiores da visão elaborada por Marr. Os captadores, quer dizer, os dispositivos de transformação dos fluxos de energia externa em mensagens materiais para uso interno, fazem além disso parte, segundo Pylyshyn, enquanto funções, da *arquitectura funcional*.

A outra proposta de organização inspira-se numa velha ideia da psicologia, a de faculdade. Num livro recente²⁶, Fodor dedica-se a reabilitar essa hipótese, na esperança de que ela permitirá explicar as dificuldades actuais e determinar aquelas que poderão um dia ser ultrapassadas. Ele propõe um modelo de aparelho cognitivo do homem que comporta três tipos de componentes:

- Captadores;
- *Input systems* ou módulos;
- Uma unidade central.

Um módulo trata exclusivamente um certo tipo de informação, estreitamente limitado, que ele recebe de certos captadores, ou talvez de outros módulos. Ele é impermeável às influências cognitivas: o resultado que produz depende apenas do fluxo dos dados; não há nenhuma interacção ou retroacção com processos situados a jusante ou em paralelo. O módulo opera de maneira obrigatória, automática e muito rápida. Entre os módulos figuram, para além daqueles, esperados, que tratam os diversos tipos de informação sensorial, aqueles que tratam certas informações linguísticas (podemos conjecturar que se trata pelo menos das informações fonológicas e sintácticas).

A unidade central, a que Fodor chama sistema de fixação da crença (*belief*), corresponde *grossa modo* às funções cognitivas superiores (com excepção, portanto, da linguagem, pelo menos em algumas das suas dimensões). Contrariamente aos módulos, esse sistema seria totalmente desprovido de estrutura, estando potencialmente cada um dos seus pontos ligado a todos os outros. Como, em suma, afirma George Miller, o psicólogo de Harvard que a este propósito cita Pylyshyn, uma característica da nossa vida mental é que, com efeito, nós podemos à vontade fazê-la tomar quase sempre não importa que caminho. Daí se segue, segundo Fodor, que o sistema de fixação da crença escapará sem dúvida sempre a uma teorização científica completa: o programa cognitivista, na sua componente puramente científica, deveria então restringir-se ao estudo dos módulos e da arquitectura do conjunto. A do sistema central, essa, continuaria impregnada

de filosofia... Estas ideias, em grande parte especulativas e programáticas, como o próprio Fodor reconhece, são naturalmente objecto das mais animadas discussões.

Mas, independentemente do que possa advir dessas hipóteses, a ideia de diferenciações modais e regionais dos fenómenos cognitivos parece abrir caminho. Diferentes categorias de fenómenos apelariam então a teorias diferentes, e funções elementares aparentemente idênticas, mas intervindo em processos diferentes, poderiam revelar-se à análise serem realizadas de uma maneira que varia radicalmente de um processo para outro. Estamos bem longe das intuições da primitiva IA!

O império da teoria

Paralelamente a estas tentativas de organização prévia do campo de estudo, e em ligação com elas, a teoria enquanto tal é objecto de reexame.

Após os excessos da modelização anárquica e das generalizações irreflectidas, é claramente sentida a necessidade de uma metodologia rigorosa. Pylyshyn e Marr, por exemplo, tentam definir regras estritas que tornem possível uma avaliação dos resultados com a ajuda de questões tais como: que problema teórico de cálculo está resolvido? A solução é fundada teoricamente ou é *ad-hoc*?, etc.

Levanta-se, em particular, uma distinção entre dois níveis *bricolage*. Que a natureza *bricole* é o seu direito e nada podemos contra isso. Mas uma *bricolage* de teoria não é uma teoria da *bricolage*. Por outras palavras, que a natureza não se deixe encerrar numa teoria uniforme não nos autoriza a adoptar sob a designação de teoria uma justaposição arbitrária de soluções. Trata-se de «separar a natureza pelas suas articulações», e as articulações de uma teoria complexa devem ser baseadas teoricamente. Roger Schank, que dirige em Yale um dos principais centros mundiais de investigação em IA, declarava, em suma, que não ficaria de modo algum decepcionado por verificar que cada novo problema de IA exige para a sua solução um método particular. Estávamos em 1978. Um biólogo, John Marshall, ficou indignado²⁷. Essa indignação parece ser partilhada daí em diante.

A mesma preocupação de rigor teórico anima empreendimentos conduzidos num quadro estritamente cognitivista mas preocupados em preservar a sua independência frente a hipóteses forçadas não motivadas ou a projectos globais demasiado precisos, cuja formulação lhes parece demasiado prematura. É o que acontece com a teoria formal da aprendizagem, que se desenvolve vigorosamente desde há uma dezena de anos, ou com a muito recente teoria da pertinência²⁸. Cada um tem a ambição de contribuir, de modo talvez decisivo, para a construção de uma teoria cognitivista geral, mas guarda-se de toda a generalização prema-

tura, o que torna possível o desenvolvimento, no modo hipotético-dedutivo, de um *corpus* muito rico de resultados sólidos, coerentes e frequentemente surpreendentes.

Surge um pensamento dos limites. Vimos como Fodor vinha excluindo da teoria o conjunto dos processos cognitivos superiores, próximos da linguagem. Pylyshyn concede que não somente os humores mas também o desenvolvimento cognitivo e a aprendizagem possam dever ser enquadrados entre os «efeitos» não cognitivos. Daí a questão, que ele levanta no final do seu *Computation and Cognition*: de que é a ciência cognitiva a ciência?

Se se verificasse que a teoria, afinal de contas, devesse efectivamente renunciar a ocupar todo ou parte do terreno que ela ambicionava à partida, que viria a ser o estado da cognição? Não será necessário desde logo considerar a ideia de um *discurso descritivo* rigoroso, dando conta de maneira tão económica e evocadora quanto possível de um conjunto de fenómenos dos quais nenhuma *teoria* parece possível? Poderia acontecer então que não se pudesse nunca falar de ciências cognitivas, a não ser no sentido de que hoje se fala de ciências humanas. Poder-se-ia assim interpretar os movimentos de vaivém que observamos nas ciências cognitivas, tal como os confrontos veementes aos quais dão lugar, como tantos outros sintomas da instabilidade talvez inerente ao projecto das ciências humanas.

Os cognitivistas, bem entendido, mesmo os mais prudentes, não vão tão longe, e uma tal conclusão parece-me prematura. Mas não será importante poder somente considerá-la e mais ainda não considerá-la como injuriosa relativamente aos investigadores que, desde há vinte e cinco ou trinta anos, se esforçam por dar forma ao projecto cognitivista?

Ciências cognitivas e ciência cognitiva

Da história do empreendimento cognitivista podem certamente ser tirados múltiplos ensinamentos. Limitar-me-ei a concluir com duas considerações. Somos, em primeiro lugar, atingidos pela intensa actividade epistemológica à qual deram lugar todos estes trabalhos. Esta actividade foi desenvolvida tanto pelos cientistas como pelos filósofos e parece ter sido útil a todos os interessados. Ela ajudou os cientistas a apreender o alcance e o quadro natural dos seus primeiros resultados, levando-os nomeadamente a reflectir sobre o sentido ambíguo de uma das suas metáforas favoritas, a do primeiro passo. Ela permitiu-lhes apreender melhor o papel que a metodologia deve desempenhar, o de relativizar a importância dos manifestos e a de apreciar a dos programas de investigação. Ela favoreceu o relacionamento das diversas aproximações, desenvolvidas por disciplinas que não têm nenhum hábito de contacto. Ela pô-las talvez

mesmo na via de certos resultados, de certos conceitos. Inversamente, ela provocou na filosofia do espírito e na epistemologia uma renovação cujo efeito ainda não deixou de se fazer sentir.

Em segundo lugar, as pesquisas cognitivas parecem caracterizadas por uma forte tendência para a interdisciplinaridade. E parece-me importante observar que o enorme trabalho realizado pelos investigadores das diversas disciplinas a ela relativas não se deixa decompor nas partes psicológica, informática, linguística, biológica, filosófica, etc., da qual se teriam encarregue respectivamente os psicólogos, os informáticos, etc.

Contudo, um estudo aprofundado dos diferentes programas em investigação cognitiva mostraria que, penso eu, a junção de várias disciplinas só foi verdadeiramente fecunda no contexto preciso de objectivos bem definidos. As ciências que hoje abordam os fenómenos cognitivos não estariam portanto, de modo nenhum, votadas à fusão. Se este diagnóstico é exacto, quaisquer que possam ser as oportunidades das ciências cognitivas, as de uma ciência cognitiva parecem escassas.

NOTAS

¹ A linguística e a doutrina chomskianas (cf. por exemplo Chomsky, 1975), por muito considerável que tenha sido o seu papel por outros motivos, não parecem ter influenciado duravelmente a corrente saída da IA (cf. Newell, 1983). Elas mereceriam de qualquer modo uma análise separada que sairia do quadro desta comunicação.

² *A physical symbol system has the necessary and sufficient means for general intelligent action.* Allen Newell & Herbert A. Simon, Tenth Turing Award Lecture, 1975; reproduzido em Haugeland (1981).

³ Herbert Simon é igualmente laureado do Prémio Nobel da Economia.

⁴ Alan M. Turing (1912-1954), matemático e logicista inglês, foi o primeiro a ter a ideia dos computadores modernos e elaborou a sua teoria matemática. Num célebre artigo publicado em 1950 (cf. Anderson, 1954) expunha as ideias principais do que dentro em pouco viria a tornar-se a inteligência artificial.

⁵ Dretske (1981) constituiu a mais estruturada defesa filosófica desta posição, que Fodor contesta.

⁶ É pouco mais ou menos nestes termos que Dennett (num artigo a publicar) apresenta o cognitivismo «ortodoxo» do qual Fodor foi durante muito tempo o mais eloquente defensor. A designação irónica que Dennett dá a esta tese (*High Church computationalism*) indica o distanciamento que um filósofo pró-cognitivista pode tomar relativamente a uma expressão maximalista da doutrina.

⁷ Ver por exemplo o artigo de Nils Nilsson no *AI Magazine* (Verão de 1984) e o debate que se seguiu na mesma revista, e Feigenbaum & McCorduck (1983).

⁸ Ver McCorduck (1979).

⁹ Numerosas funções calculáveis não são num tempo razoável por uma máquina mesmo futurista, submetida a limitações tais como a velocidade da luz e as dimensões do universo.

¹⁰ *Unpopular Essays*; citado em Pylyshyn (1984), p. 257.

¹¹ Ver em Johnson-Laird & Wason (1977) uma passagem consagrada ao matemático A. C. Aitken extraída de um artigo de I. M. L. Hunter (1966).

¹² Estes termos, propostos por Pylyshyn, parecem-me preferíveis aos de *representação* e de *simulação* utilizados no mesmo contexto para diversos outros fins.

¹³ Os sistemas periciais estão no entanto longe de ser perfeitos. Ver Dreyfus & Dreyfus (1986), mas também os artigos de Randall Davies e de John Doyle no *AI Magazine* (1982, 1984) para interessantes reflexões críticas vindas de especialistas desses sistemas.

¹⁴ Ver por exemplo Johnson-Laird & Wason (1977) e o artigo panorâmico de John Anderson na *Artificial Intelligence*, 23 (1984).

¹⁵ Na introdução a Fodor (1981).

¹⁶ Com excepção dos estados puramente «qualitativos», como os que correspondem à experiência de uma dor.

¹⁷ Sobre esta noção ver Fodor (1981) e Dennett (1978).

¹⁸ O que Cummins (1983, pp. 54, 84 e segs.) chama a «*-cognição».

¹⁹ Foi Dreyfus (1972-1979) que abriu a via da crítica fenomenológica da IA; apoiou-se nomeadamente em Heidegger, Merleau-Ponty e Wittgenstein. Para a possibilidade de uma verdadeira ciência cognitiva que corresponda às ambições de Husserl, cf. Dreyfus (1982).

²⁰ A tentativa mais aprofundada é a de Lucas [ver artigo de 1961 reproduzido em Anderson (1964), e Lucas (1968)]. Alguns contributos essenciais são devidos a Benacerraf (1967) e a Lewis (1969 e 1979). Ver a opinião do próprio Gödel em Wang (1974). Ver também Dennett (1978). A obra fundamental sobre a questão é ao que a ela está ligado é Webb (1980).

²¹ Ver Block (1983), a obra colectiva Block (1981), Kosslyn (1980) e Pylyshyn (1984).

²² Sobre a delicada distinção entre numérico (*digital* em inglês) e analógico, ver Goodman (1968), Lewis (1971), Haugeland em Biro & Shahan (1982), Pylyshyn (1984).

²³ Ver Varela (1979).

²⁴ Ver McCulloch & Pitts (1943), Von Neumann (1966), Minsky & Papert (1969), Atlan (1972) (1979), os artigos de Atlan Weisbuch e Fogelman Soulié em Dumouchel & Dupuy (1983), Hinton & Anderson (1981), Hopfield (1982), Kohonen (1984), McClelland & Rumelhart (a publicar).

²⁵ Ver Pylyshyn (1984).

²⁶ Fodor (1983).

²⁷ Em Caplan (1982).

²⁸ Ver Osherson, Staub & Weinstein (1986) e Sperber & Wilson (a publicar).

BIBLIOGRAFIA

- A. R. Anderson, ed., *Minds and Machines*, Prentice Hall, 1964.
H. Atlan, *L'organisation biologique et la théorie de l'information*, Hermann, 1972.
H. Atlan, *Entre le cristal et la fumée*, Seuil, 1979.
P. Benacerraf, «God, the Devil, and Gödel», *Monist*, LI, 1967.
J. I. Biro e R. W. Shahan, ed., *Mind, Brain, and Function*, University of Oklahoma Press, 1982.
N. Block, ed., *Imagery*, Bradford Books/MIT Press, 1981.
N. Block, «Mental pictures and cognitive sciences», *Phil. Rev.*, XCII, 1983.
D. Caplan, *Biological Studies of Mental Processes*, MIT Press, 1980.
N. Chomsky, *Reflections on Language*, Pantheon Books, 1975.
R. Cummins, *The Nature of Psychological Explanation*, Bradford Books/MIT Press, 1983.
D. Davidson, «Mental Events», in L. Foster e J. W. Swanson, ed., *Experience and Theory*, University of Massachusetts Press, 1970.
D. Davison, «The Material Minds», 1973, in Haugeland, 1981.
D. C. Dennett, *Brainstorms*, Bradford Books & Harvester Press, 1978.

- D. C. Dennett, «The Logical Geography of Computational Approaches: a View from the East Pole», MIT Sloan Conference, 1984.
- D. C. Dennett, «Cognitive Wheels: the Frame Problem of AI», in C. Hookway, ed., *Minds, Machines and Evolution*, Cambridge University Press, 1985.
- F. I. Dretske, *Knowledge and the Flow of Information*, Bradford Books/MIT Press, 1981.
- H. L. Dreyfus, *What Computers can't Do*, Harper Colophon Books/Harper and Row, 1972-1979.
- H. L. Dreyfus, ed. com H. Hall, *Husserl, Intentionality and Cognitive Science*, Bradford Books/MIT Press, 1982.
- H. Dreyfus et S. Dreyfus, *Mind over Machine*, MacMillan/The Free Press, 1986.
- P. Dumouchel e J.-P. Dupuy, dir., *L'auto-organisation*, colloquio de Cerisy, Seuil, 1983.
- E. A. Feigenbaum e P. McCorduck, *The Fifth Generation*, 1983, Addison-Wesley.
- F. Flores e T. Winograd, *Understanding Computers and Cognition*, Ablex Publ.
- J. A. Fodor, *The Language of Thought*, Crowell & Harvard University Press, 1975.
- J. A. Fodor, *Representations*, Bradford Books/MIT Press, 1980.
- J. A. Fodor, *The Modularity of Mind: An Essay on Faculty Psychology*, Bradford Books/MIT Press, 1983.
- N. Goodman, *Language of Art*, Bobbs-Merrill, 1968.
- J. Haugeland, ed., *Mind Design*, Bradford Books/MIT Press, 1981.
- J. Haugeland, *Artificial Intelligence: The Very Idea*, Bradford Books/MIT Press.
- G. E. Hinton e J. A. Anderson, *Parallel Models of Associative Memory*, Erlbaum, 1981.
- A. Hodges, *Alan Turing: The Enigma*, Burnett, 1983.
- D. R. Hofstadter, *Metamagical Themas*, Basic Books, 1985.
- J. J. Hopfield, «Neural networks and physical systems with emergent collective computational abilities», *Proc. Nat. Acad. Sci. USA*, 79, 1982, pp. 2554 e segs.
- P. N. Johnson-Laird e P. C. Wason, *Thinking: Readings in Cognitive Science*, Cambridge University Press, 1977.
- T. Kohonen, *Self-Organization and Associative Memory*, Springer Verlag, 1984.
- S. Kosslyn, *Image and Mind*, Harvard University Press, 1980.
- D. Lewis, «Lucas Against Mechanism», *Philosophy*, 44, 1969.
- D. Lewis, «Analog and Digital», *Nous*, V, 1971.
- D. Lewis, «Lucas Against Mechanism II», *Canadian J. of Phil.*, IX, 3, 1979.
- J. R. Lucas, «Minds, Machines and Gödel», *Philosophy*, 36, 1961.
- J. R. Lucas, «Satan Sullified: A Rejoinder to Paul Benacerraf», *Monist*, 52, 1968.
- J. L. McClelland e D. E. Rumelhart, ed., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*.
- P. McCorduck, *Machines Who Think*, Freeman, 1979.
- W. S. McCulloch e W. Pitts, «A Logical Calculus of the Ideas Immanent in Nervous Activity», *Bull. Math. Biophysics*, 5, 1948.
- D. Marr, *Vision*, Freeman, 1982.
- M. Minsky, *Semantic Information Processing*, MIT Press, 1968.
- M. Minsky e S. Papert, *Perceptrons: An Introduction to Computational Geometry*, MIT Press, 1969.
- J. von Neumann, *The Theory of Self-Reproducing Automata*, University of Illinois Press, 1966.
- A. Newell, «Intellectual issues in the history of artificial intelligence», in F. Machlup e U. Mansfield, ed., *The Study of Information: Interdisciplinary Messages*, Wiley, 1983.
- D. N. Osherson, M. Stob e S. Weinstein, *Systems That Learn*, Bradford Books/MIT Press, 1986.
- H. Putnam, *Mind, Language and Reality — Philosophical Papers*, vol. 2, Cambridge University Press, 1975.
- Z. W. Pylyshyn, *Computation and Cognition*, Bradford Books/MIT Press, 1984.
- G. Ryle, *The Concept of Mind*, 1949, Hutchinson. Reimpr. Penguin, 1980.
- R. C. Schank e R. P. Abelson, *Scripts, Plans, Goals and Understanding*, Erlbaum, 1977.
- J. R. Searle, 1980, «Minds, Brains, and Programs», *Behavioral & Brain Sci.*, 3, 1980. E. in Haugeland, 1981.
- B. Shanon, «Remarks on The Modularity of Minds», Goldie Rotman Center working paper n.º 13, The Hebrew University, Jerusalém, 1985.
- B. Shanon, «Reflections on the limitations of semantic representations».
- H. A. Simon, *The Sciences of the Artificial*, 1969/1981, MIT Press.
- D. Sperber e D. Wilson, *Relevance: Communication and Cognition*, Blackwells.
- F. J. Varela, *Principles of Biological Autonomy*, North Holland, 1979.
- H. Wang, *From Mathematics to Philosophy*, Routledge & Kegan Paul, 1974.
- P. Watzlawick, *The Invented Reality*, W. W. Norton, 1984.
- J. C. Webb, *Mechanism, Mentalism, and Metamathematics*, Reidel, 1980.