UNIVERSALIDADE E CIÊNCIAS COGNITIVAS

António Machuco Rosa

Universidade Lusófona de Humanidades e Tecnologias – Departamento de Ciências da Comunicação e da Informação

Neste artigo apresentar-se-ão algumas linhas dominantes dos principais paradigmas de investigação contemporânea em física e ciências cognitivas. Se não nos ocuparemos aqui do conceito sociológico de paradigma, estará contudo implícito ao longo do texto que ele pode ser tomado na acepção, científica, de "classe de universalidade". É o conceito de universalidade que constituirá o núcleo do artigo. Ele aponta na direcção de um novo paradigma interdisciplinar que teve origem na física dos fenómenos críticos e está a ser traduzido no campo das ciências cognitivas. O conceito de universalidade, possibilitando uma concepção rigorosa dos fenómenos de auto-organização e de emergência, permite desenvolver uma filosofia geral que no texto é designada por funcionalismo dinâmico. Algumas das principais doutrinas científico-filosóficas contemporâneas serão então agrupadas em torno dos diversos funcionalismos propostos durante a segunda metade deste século. Em particular, sustentar-se-á que a natureza e estatuto das ciências cognitivas contemporâneas se torna inteligível a partir da oposição entre funcionalismo simbólico e funcionalismo dinâmico.

I

Durante os anos quarenta, surgiu um movimento de pensamento onde se encontram em germe a totalidade das teorias que hoje em dia se afrontam no terreno multidisciplinar designado por "ciências cognitivas". Tratava-se da Cibernética. Seguindo a obra dos seus principais teorizadores, W. McCulloch, W. Pitts, N. Wiener, entre outros, é possível ver como a física, por um lado, e a lógica formal, por outro, são encaradas como duas alternativas para a modelização do cérebro (*cf.* o excelente Dupuy, 1994). Essa competição perdurou até hoje, e este artigo visa dar disso conta. Contudo, nos anos cinquenta, os modelos inspirados na lógica formal foram adquirindo primazia, facto em parte explicável por ser a lógica que primeiro apurou um conceito de universalidade cujas possibilidades de aplicação ao estudo da cognição pareciam evidentes.

Existe uma certa acepção do conceito de universalidade no âmago da definição de "lógica". Recorde-se que, em lógica formal, se diz que uma fórmula é universalmente válida se ela é válida em qualquer interpretação das suas variáveis proposicionais e variáveis de indivíduo. Noutros termos, uma fórmula é universalmente válida se ela é válida em qualquer seu modelo. Constituiu um acontecimento historicamente da maior importância ter sido mostrado, por Gödel em 1929, que essa noção de validade se precisava em função do chamado teorema de completude. Esse teorema estabelece uma equivalência entre uma exposição sintáctica e uma representação semântica de certos sistemas de lógica. Mais precisamente, o chamado teorema da completude semântica diz que se uma certa fórmula é uma consequência semântica de um certo conjunto de proposições prévias, então ela é igualmente uma consequência sintáctica dessas mesmas proposições. Essa propriedade é uma propriedade dos

sistemas lógicos à primeira ordem: nesses sistemas, qualquer proposição verdadeira é igualmente uma proposição formalmente demonstrável (pode consultar-se qualquer manual de lógica para a demonstração desse teorema; *cf.* em particular o clássico Kleene, 1987). Esta coincidência entre validade universal e demonstração sintáctica é aqui importante, visto ela se encontrar na base de algumas das ideias que surgiram em ciências cognitivas. Ela aponta para o que aqui chamamos um invariante de todos os invariantes. Por tal expressão entendemos inicialmente (o seu âmbito alargar-se-á mais abaixo) o facto de a "lógica" ser um invariante por relação a qualquer tipo de notação escolhida. Nessa medida, uma tal expressão mais não é que um corolário do teorema da completude de Gödel. Ela é uma glosa da palavra "lógica".

É no entanto sobejamente conhecido que a propriedade de completude não é generalizável a qualquer tipo de notação escolhida. Deve repetir-se que as asserções anteriores apenas são exactas no que respeita à lógica à primeira ordem (cálculo de predicados com interdição de quantificar sobre as letras de predicados). Para sistemas à segunda ordem, tem-se o famoso teorema da incompletude de Gödel: em sistemas suficientemente ricos para exprimir a totalidade da aritmética (o sistema de Peano, por exemplo), existe um resíduo inilimináveI de objectos, isto é, existem fórmulas semanticamente verdadeiras mas que não são sintacticamente demonstráveis. Esta referência ao teorema da incompletude de Gödel não visa estabelecer as bases para demonstrar teoricamente as insuficiências dos modelos cognitivos baseados nas linguagem simbólicas formais; fazer isso seria seguir a linha de Penrose (Penrose, 1989). Referimos esse teorema por a sua demonstração ter exigido de Gödel a invenção de uma engenhosa técnica de codificação dos símbolos formais na aritmética: existe uma função que permite representar qualquer proposição da lógica simbólica como uma proposição da aritmética. Essa técnica foi decisiva para o surgimento de um conceito destinado a desempenhar um papel fundamental nas emergentes ciências cognitivas: o conceito de máquina de Turing (*cf.* Hodges, 1983). É esse conceito que nos permite tomar "universalidade" numa nova acepção.

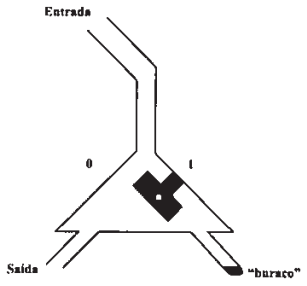
Recordemos rapidamente que A. Turing imaginou uma máquina que poderia encontrar-se em certos estados internos (os estados actuais da máquina) e que possuiria uma fita ilimitada nos dois sentidos, fita essa divida em células. Um certo símbolo estaria (ou não) escrito em cada uma das células. Além disso, a máquina possuiria uma cabeça de leitura permitindo realizar as seguintes operações: *a*) verificar se a célula sob leitura possui ou não um símbolo: *b*) apagar o símbolo no caso de ele existir; *c*) inserir um símbolo no caso de não haver nenhum; *d*) após qualquer uma das duas operações anteriores, a máquina pode deslocar a sua cabeça de leitura ao longo das duas direcções da fita. Note-se que este dispositivo realiza apenas operações elementares (apagar, inserir, mover-se um passo para a esquerda ou direita). Ela representa o arquétipo do pensamento discreto.

Através de uma codificação apropriada, uma máquina de Turing particular pode calcular uma certa função numérica também particular. O passo decisivo de A. Turing consistiu em demonstrar a existência de uma certa máquina de Turing universal. Esta é capaz de simular não importa que máquina de Turing particular capaz de calcular uma função numérica também particular. Basta para isso que a máquina universal tenha escrito na sua fita um símbolo que codifique o funcionamento da máquina a simular. A máquina universal simula então o processo executado pela máquina simulada. Este é um dos sentidos de universalidade: uma máquina de Turing que simula qualquer função numérica calculável de modo efectivo[1]. Especificamente, ela pode simular as funções lógicas que, como atrás se referiu, os trabalhos de Gödel mostraram poderem ser representadas por funções numéricas. Do ponto da modelização cognitiva, o conceito de máquina de Turing abriu uma perspectiva de uma enorme riqueza: se uma certa função cognitiva pode ser simulada por uma certa máquina de Turing (por um certo programa), existe então uma máquina universal capaz de simular ***qualquer*** função cognitiva.

Os símbolos da máquina de Turing podem ser apenas dois: 0 e 1. Isso permite-nos compreender a passagem a máquinas reais desse artefacto conceptual que é a máquina de Turing. Para o fazermos de modo cabal teríamos de introduzir o conceito de máquina de von Neumann, base dos modernos computadores digitais. Mas para os objectivos que temos em vista não é necessário ir tão longe. Basta-nos apresentar o diagrama de um sistema físico bastante elementar, capaz no entanto de calcular uma certa função numérica. Verifique-se que a tabela seguinte representa a tabuada da adição em notação binária:

Entrada	Estado inicial	Estado final	Saída
0	0	0	0
0	1	1	0
1	0	1	0
1	1	0	1

Note-se então que um dispositivo como o seguinte implementa a adição:



[1] "Efectivo" designa o conjunto de funções que, num sentido intuitivo, dizemos serem as funções calculáveis. A conjectura de Turing consistiu em dizer que essas funções são as recursivas gerais, isto é, as funções do tipo a+b, a.b, a função sucessor, etc.

onde o T invertido designa a função de transição + entre dois estados, 0 e 1, o mecanismo sendo posto em acção por uma bola que entra (ou não) pela Entrada. Assim, por exemplo, se entra uma bola, e se o estado inicial é 1 (como sucede no diagrama), o dispositivo passa para o estado final 0 (o torniquete roda) e a bola vai pela saída ("vai um", em notação binária).

É evidente que esta máquina não possui memória, pelo que as suas *performances* estão limitadas ao máximo. Ela ilustra contudo um ponto de alcance bastante geral, base daquilo a que se chama o *paradigma simbólico* em ciências cognitivas. Um cognitivista como Ned Block pô-lo em destaque (Block, 1991). Embora seja possível construir uma infinidade de mecanismos "calculadores" com materiais físicos e com uma arquitectura bastante diferente da acima apresentada, pode-se contudo mostrar que todos eles são equivalentes do ponto de vista da função computada. Block apresenta diferentes exemplos de mecanismos susceptíveis de implementar identica o operador lógico "e", salientando que qualquer um deles preenche a função requerida desde que tenhamos dois, e apenas dois, estados, assim como algo representando a transição entre esses estados. Sempre de acordo com Block, esse ponto seria fundamental para apreciar a natureza da modelização cognitiva que utiliza linguagens de tipo simbólico. Essas linguagens são universalmente realizáveis, no sentido de poderem ser implementadas em qualquer mecanismo que tenha uma arquitectura minimamente adequada. No fundo, isso não nos deve surpreender, pois a lógica formal (e sobretudo uma lógica de tipo booleano) é precisamente um invariante de todos os invariantes. Ela é praticamente indiferente a qualquer especificação que os objectos possam possuir: representa o objecto (o objecto em geral, em linguagem filosófica) na sua máxima generalidade. A lógica simbólica será pois universalmente representável.

É em torno deste conceito de universalidade (de que os computadores digitais seriam o paradigma) que se estabeleceu o paradigma simbólico – até à bem pouco tempo dominante em ciências cognitivas. Como Putnam (Putnam, 1960) mostrou, o paradigma simbólico é um funcionalismo simbólico. Trata-se de um funcionalismo na medida que se supõe a existência de um nível autónomo de realidade descrito pelas linguagens simbólico-formais, nível esse independente da estrutura física particular em que ele se possa implementar. Independente da sua implementação, esse nível é no entanto universal no sentido de se poder implementar não importa em que realidade material. Faz-se então a hipótese que esse nível corresponde ao nível cognitivo. Nasce assim a ideia segundo a qual a cognição é descrita por uma linguagem simbólica implementável num material físico (ou biológico, visto os materiais de implementação constituirem uma classe de equivalência por relação à estrutura simbólica). A cognição é um sistema físico-simbólico

O funcionalismo simbólico foi desenvolvido com grande precisão por J. Fodor (*cf.* Fodor, 1975; 1983; *cf.* também Pylyshyn, 1984). Resumidamente, Fodor propõe uma Teoria Representacional da Mente segundo a qual os estados intencionais (crenças, desejos, etc.) possuem uma relação com certas proposições que constituem uma "linguagem mental" interna. Essa linguagem mental possuirá propriedades análogas às das linguagens lógicas: existirão estados mentais que são as instâncias (os *tokens*) dos tipos (*types*) invariantes dessa linguagem (analogamente à relação instância/tipo presente na dualidade sintaxe/semântica existente em lógica formal). Os processos mentais serão então relações entre as instâncias das representações mentais, relações que são ainda uma instanciação das relações sintácticas que ligam os diferentes tipos da linguagem formal interna. A essa Teoria Representacional, Fodor acrescenta a Teoria Computacional, segundo a qual as relações sintácticas estabelecendo as ligações entre os estados mentais são relações computacionais. Essas ligações são apenas determinadas pela estrutura sintáctica da linguagem, e não pelo seu conteúdo semântico ou pela natureza dos processos físicos em que elas são supostas implementarem-se (*cf.* um resumo em Lower & Rey, 1991).

O funcionalismo simbólico salienta a autonomia do nível funcional por relação ao nível físico (e biológico). No entanto, parecem surgir certas ambiguidades a partir do momento em que os funcionalistas afirmam que a cognição é um sistema físico-simbólico. A expressão "físico" é, neste contexto, geradora de uma espécie de ilusão. De facto, mesmo considerando apenas o exemplo elementar da máquina acima apresentada, pode constatar-se que o simbólico (a tabuada da adição) que se implementa não tem nada a ver com a física propriamente dita da máquina, se por "física" entendermos aquilo que surge nos livros usuais acerca dessa disciplina. A Física não é uma linguagem simbólica. O nível simbólico é completamente distinto do nível propriamente físico da máquina (em nossa opinião, J. Searle, 1992, argumentou correctamente acerca desse ponto). A noção de "realização universal" não é algo que emerge a partir da física. Ela é um dado *a priori*, no sentido de determinar à partida a descrição que podemos fazer acerca do comportamento da máquina. Para que o nível físico adquira sentido semântico tem de existir uma interpretação *exterior* ao funcionamento do dispositivo mecânico. A descrição simbólica do funcionamento da máquina em parte alguma alude a "velocidade", a "mudanças contínuas no tempo", isto é, às equações do movimento e à sua integração. Na expressão que H. Pattee foi buscar à física teórica (Pattee, 1997), as linguagens simbólicas são incoerentes, o que significa que elas são completamente arbitrárias por relação à localização e escalas espácio-temporais; a universalidade simbólica é conseguida precisamente devido a essa arbitrariedade.

Deve aqui observar-se que o estatuto das linguagens simbólicas é o mesmo que elas possuem numa teoria matemática axiomatizada. Numa perspectiva axiomática acerca da validade de uma teoria matemática, pode afirmar-se que a validade universal está assegurada na medida em que os símbolos e as regras que transformam as proposições da teoria são introduzidas por nós e definem *a priori* os objectos e propriedades dessa mesma teoria; os símbolos e regras são universais por sermos nós a introduzir essa universalidade (Machuco Rosa, 1993). Noutros termos, o

controlo sobre a teoria é elevado – mas não absoluto devido aos teoremas de incompletude – por se determinar à partida quais as propriedades em relação às quais os símbolos e regras nada dizem. É isso que se verifica na implementação do simbólico: asseguramos à partida – *constringindo* fortemente o material físico – que a única coisa relevante para o sistema sejam as transições sintácticas entre os símbolos; tudo o resto (por exemplo, a dissipação de energia na máquina) é contingente (indeterminado) por relação à linguagem simbólica previamente definida. Por definição, as linguagens simbólicas são independentes de uma implementação particular, e são universais no mesmo sentido em que se diz que uma proposição matemática é universal e necessária: esta também é independente dos materiais físicos ou dispositivos neuronais utilizados para a demonstrar. A conclusão geral só pode ser que a "realização universal" não é algo que emerge a partir da física. Ela é um dado *a priori*, no sentido de determinar ou constringir à partida a descrição que podemos fazer acerca do comportamento de uma máquina simbólica universal. *O nível simbólico é apenas um instrumento descritivo*, imposto *a priori* a uma realidade que lhe pode ser eventualmente irredutível

Nessas condições, melhor seria que os funcionalistas deixassem de utilizar a expressão "física", tirando explicitamente a consequência que por vezes eles parecem evitar: o funcionalismo simbólico representa uma forma extrema de dualismo. Trata-se precisamente de um dualismo entre físico e simbólico. A cognição, o sentido, etc., são irredutíveis à física (como Fodor não deixou de frequentemente sublinhar). Esse dualismo é um dos principais problemas teóricos do funcionalismo simbólico, dele decorrendo o completo mistério que rodeia a implementação das linguagens simbólicas em cérebros reais (e não em computadores)[2]. A questão reside então em saber se, precisando um pouco o que se entende por "física" – mostrando em particular que "física" não designa necessariamente a "física" que os funcionalistas têm em vista quando falam dessa disciplina –, não será possível desenvolver o quadro teórico susceptível de reduzir o dualismo absoluto presente no funcionalismo clássico. Para o fazer, não se deverá partir das linguagens simbólicas como um dado *a priori* constitutivo, mas sim de um certo nível da física, vendo como se encontra aí uma nova acepção de universalidade que permite começar a pensar o nível simbólico como um fenómeno de emergência.

II

Embora a possibilidade de naturalizar as estruturas simbólicas já se encontrasse em germe nos trabalhos de alguns dos fundadores do movimento cibernético (era em particular o caso de W. McCulloch), é apenas devido a um conjunto de teorias físicas surgidas nos últimos vinte anos que uma aproximação dos fenómenos cognitivos em termos de emergência se começou a desenvolver em bases suficientemente claras. A "física" a que aqui se alude não é a física das partículas elementares, nem tão pouco a teoria física do universo considerado como um todo. Trata-se da *teoria geral dos fenómenos críticos*. Partindo das teorias clássicas de Landau, essa teoria atingiu, graças aos trabalhos de L. Kadanoff, B. Widon e K. Wilson, um estado quase perfeito. Ela tornou-se então um verdadeiro quadro transdisciplinar. Aplicada aos fenómenos cognitivos, permite novos tipos de funcionalismos. Apesar de se tratar de uma teoria bastante sofisticada do ponto de vista matemático, é possível apresentar aqui de modo resumido as suas ideias essenciais.

Um exemplo da teoria dos fenómenos críticos é fornecido pelos modelos dos sistemas de Ising. Tratam-se de modelos de sistemas magnéticos segundo os quais os átomos possuem um momento magnético ou *spin*, podendo suceder que esse *spin* possa apontar apenas em duas direcções opostas do espaço. O sistema de *spins* mais simples é o sistema de *spins* unidimensional. A energia da interacção magnética entre os elementos do sistema pode ser representada pelo Hamiltoniano **H**:

$$1)\ \mathbf{H} = K \sum_{ij} s_i s_j$$

onde K é o parâmetro de acoplamento entre dois *spins*, e s designa os estados {0,1} que o *spin* pode assumir. No caso em que K tem valor positivo, a energia total dada pelo Hamiltoniano minimiza-se quando os *spins* apontam todos na mesma direcção. Fala-se, nesse caso em interacção ferromagnética (a interacção antiferromagnética correspondendo ao caso em que K é negativo e os *spins* adjacentes apontam alternadamente nas duas direcções): existe então um estado macroscópico global (os *spins* apontam todos numa mesma direcção, no caso ferromagnético). Ora, sucede que o comportamento desse sistema pode depender crucialmente de um parâmetro determinando as suas transições. Esse parâmetro pode ser K, ou então uma temperatura, T, sendo exactamente em $T=0$ que o sistema atinge um mínimo de energia correspondendo ao estado ferromagnético de ordem macroscópica global. Pode pois afirmar-se que existe em $T=0$ uma espécie de transição de fase, isto é, o ponto fixo $T=0$ pode ser encarado como um ponto crítico. No caso do sistema de Ising unidimensional, apenas existe um outro ponto fixo quando $T=\infty$, ponto correspondendo ao estado de desordem (maior entropia) onde os *spins* estão dispostos aleatoriamente com idêntica probabilidade (Fisher, 1983: 50).

Retomaremos mais adiante o sistema de Ising unidimensional no quadro da teoria das redes neuronais. De momento, importa referir que esse tipo de sistema não exibe a generalidade das propriedades que são objecto da teoria dos fenómenos críticos. Esta apenas se torna paradigmática quando se consideram sistemas onde a dimensão ≥ 2. Nesses sistemas, existe uma temperatura crítica $0<T_c<\infty$ onde se dá uma transição de fase. É a explicação desse tipo de fenómenos que é objecto da teoria, dita "clássica", de Landau (para além do já referido Fisher, 1983, seguimos aqui Lopes, 1988).

[2] Naturalemente que existe um vasto conjunto de outras críticas que se podem levantar contra o funcionalismo simbolico e o seu corolário, a Inteligência Artificial Clássica. Algumas delas são referidas em A. Machuco Rosa, 1995.

A teoria de Landau visava explicar as transições de fase à segunda ordem do tipo das observáveis nos fenómenos de magnetização de sistemas como os de Ising (transições do tipo ferromagnético/paramagnético) De acordo com a termodinâmica clássica, existe nesse sistema magnético uma tendência para um estado de máxima desordem. Mas, como atrás se referiu, pode existir igualmente a emergência de um estado macroscópico de ordem global (alinhamento dos *spins*) Existe pois uma competição entre essas duas tendências, competição que representa uma ruptura da simetria presente no estado mais desordenado (o estado mais provável); essa ruptura de simetria significa a emergência de um estado mais ordenado. As rupturas de simetria características das transições de fase podem ser descritas através de um parâmetro de ordem, como, por exemplo, a magnetização nos sistemas de *spins*. A energia (livre) do sistema é então uma função desse parâmetro de ordem, e a hipótese essencial da teoria de Landau consiste em afirmar que essa função se pode desenvolver em série de Taylor em torno do ponto crítico de transição de fase. Noutros termos, supõe-se que essa função é diferenciável e analítica, o que implica negligenciar as flutuações que se dão no ponto crítico.

O problema dessa teoria clássica consiste precisamente em negligenciar o aspecto fundamental de uma teoria dos fenómenos críticos: a análise do comportamento do sistema no próprio ponto crítico, ponto usualmente controlado por uma temperatura crítica T_c. Ora, no ponto crítico, três factos são de importância fundamental:

I – No ponto crítico, a função da energia não é uma função diferenciável, portanto não exprimível analiticamente.

II – No ponto crítico, a função de correlação, ξ, que mede as interacções entre os elementos do sistema, *diverge*. Ela torna-se infinita no ponto de crítico. Isso significa que o sistema deixa de possuir uma escala característica, passando a existir invariância de escala (isto é, existe uma homotetia interna gerando uma estrutura auto-similar de tipo fractal). No ponto crítico tem-se pois um estado macroscópico de ordem global: os elementos situados a distâncias muitos superiores ao alcance das forças intra-atómicas "sentem-se" uns aos outros. No ponto crítico, apenas o comportamento macroscópico se torna relevante.

III – O comportamento crítico do sistema é essencialmente determinado pelos chamados expoentes críticos. Na realidade, constatou-se que as propriedades macroscópicas do sistema variam como uma potência simples da diferença da temperatura em relação ao respectivo valor crítico, $\Delta T = T - T_c$. O ponto crucial da teoria dos fenómenos críticos consiste em notar que se o valor da temperatura crítica T_c varia consoante o sistema físico particular, já o expoente crítico (que dá a razão da aproximação ao ponto crítico) é *universal*. Os expoentes críticos são largamente independentes de qualquer sistema físico particular, apenas dependendo da dimensão e dos graus de liberdade do parâmetro de ordem do sistema. Portanto, todos os sistemas que possuam a mesma dimensão e os mesmos graus de liberdade do parâmetro de ordem possuirão um expoente crítico com o mesmo valor. Todos os tipos de fenómenos críticos podem assim ser classificados em classes de universalidade. O conceito de universalidade designa o facto dos comportamentos críticos apenas dependerem de propriedades geométricas muito gerais, e não dos detalhes físicos e microscópicos do sistema. Os comportamentos críticos possuem *invariantes* fundamentais. Ora, a teoria de Landau fornecia valores inexactos para esses expoentes críticos.

As insuficiências da teoria de Landau obrigaram a desenvolver uma aproximação mais geral aos fenómenos críticos, permitindo determinar o valor exacto e a origem dos expoentes críticos. Essa aproximação mais geral designa-se por Grupo de Renormalização (G. R.), tendo valido a K. Wilson o prémio Nobel da Física. Resumamos as suas ideias essenciais.

Aplicado a sistemas de *spins* como o modelo de Ising bidimensional, o método do G. R. consiste, não em procurar fazer a soma (enorme) dos N *spins*, mas em dividir a rede de *spins* em blocos mais pequenos. Toma-se então um dos blocos em que a rede inicial foi subdividida e faz-se a média dos *spins* desse bloco, após o que se substitui o bloco por um único *spin* representativo da média do bloco. Essa operação implica uma considerável redução dos graus de liberdade do sistema e um aumento da distância entre os *spins* da rede[3]. Em geral, e para um sistema a d dimensões, a redução no número inicial N de *spins* é: $N \rightarrow N' = N/b^d$ [4]. De seguida, após os blocos terem sido substituídos por um *spin* representativo da média de cada bloco, a rede é reescalada por relação à rede original, fazendo agora diminuir a distância entre os spins, de acordo com a transformação $l \rightarrow l' = l/b$, onde l é a distância inicial entre os *spins*. Essas operações de redução e reescalamento podem ser representadas por um novo hamiltoniano H, com $H = R_b(H)$ onde b é o parâmetro de reescalamento espacial associado à operação de renormalização R. É o conjunto de transformações representadas por R_b ($b \geq 1$) que se designa por G. R.

Referiu-se que o resultado das transformações de renormalização consiste em criar uma rede na qual a distância entre os *spins* é b vezes maior que na rede original, após o que o reescalamento faz a distância voltar a diminuir. A operação tem como efeito eliminar as flutuações que se dão a um nível mais "microscópico": os pequenos pormenores ficam irreconhecíveis mas os maiores não são afectados. Cada transformação elimina todas as flutuações que se dão à escala da transformação, deixando presente apenas aquelas que se verificam numa escala maior, após o que se eliminam as flutuações a esta escala, e assim sucessivamente. Essa eliminação sucessiva das flutuações vai levar a que apenas as flutuações macroscópicas permaneçam. Mais exactamente, permanecem as propriedades macroscópicas de longo alcance onde a função de correlação ξ é infinita. Ora, como ξ diverge no ponto crítico, e como aí $\xi/b = \xi$, tem-se que i (ou β) é um ponto fixo de R_b. Sublinha-se de novo: a obtenção de propriedades macroscópi-

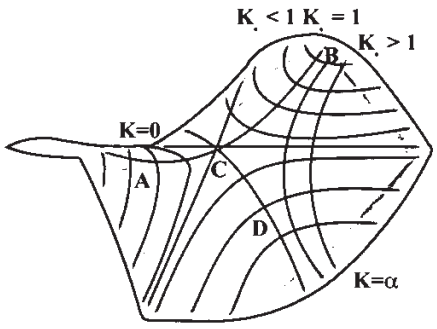
[3] A operação descrita de forma intuitiva equivale a definir um novo Hamiltoniano (na realidade, trata-se de um espaço funcional de Hamiltonianos) correspondendo ao acima escrito em (1): $H = K' S_{ij} s_i' s_j'$

[4] Se, por exemplo, tivermos uma rede bidimensional de blocos cada um de lado b=5, temos 25 *spins* em cada bloco e 250 na rede, e logo a redução é N'=250/25=10.

Humani Nihil Alienum

cas invariantes implica sempre do sistema uma redução, a redução do número (enorme) de graus de liberdade.

A iteração das transformações realizadas sobre a rede faz igualmente variar os acoplamentos, K, entre os *spins*, verificando-se a relação recursiva (iterativa): $K'=r(K)$. Por exemplo, se um novo parâmetro é K', e o anterior K, então $K'=K^2$. Se iterarmos os valores do parâmetro K (inverso da temperatura, recorde-se), o sistema pode divergir para o ponto fixo (estável) 0 (se partirmos de um valor inteiro de K superior a 1) ou para o ponto fixo ∞ (se partirmos de um valor fraccionário de K), pontos que correspondem aos pontos fixos referidos a propósito do sistema de Ising unidimensional. Mas já existe um valor inicial de K, $K=1$, em que o sistema não se altera por iteração da operação de renormalização: trata-se de um ponto fixo correspondente ao ponto crítico. Só que esse novo ponto fixo é instável, a iteração fazendo com que o parâmetro K se afaste dele (em direcção aos pontos fixos estáveis 0 ou ∞). Para o caso de um sistema de Ising bidimensional, esse comportamento é descrito por uma superfície de tipo sela (figura in Lopes, 1988).



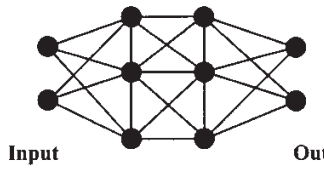
O ponto onde as curvas AB e CD se cruzam é um ponto fixo instável, representando o ponto crítico. Se $K=1$, a trajectória (o Hamiltoniano) representando o estado do sistema permanece em AB e converge para o ponto fixo instável (onde a função de correlação se torna infinita). Mas já para valores ligeiramente diferentes do parâmetro de acoplamento K, as operações de renormalização levam o sistema a aproximar-se do ponto crítico e de seguida a divergir para $K=0$ (com $K_0<1$) ou para $K=\infty$ (com $K_0>1$). É pois a estrutura geométrica da superfície (o seu declive) que determina a totalidade do comportamento do sistema. Dado o declive, sabe-se de imediato como as propriedades do sistema, controlado por parâmetros como o acoplamento K ou a temperatura, variam. Em particular, como os expoentes críticos determinam como o sistema varia ao aproximar-se do ponto crítico, esses expoentes apenas dependem do declive da superfície controlada pelos parâmetros. Mais, qualquer sistema que tenha como atractor um mesmo ponto fixo crítico possuirá os mesmos expoentes críticos, a especificidade de cada sistema físico correspondendo à posição que a sua trajectória ocupa na região de atracção do ponto fixo. Em suma, a universalidade dos expoentes críticos corresponde ao ponto fixo instável. Temos aqui um princípio de classificação geométrico intrínseco e completamente diferente das classificações lógico-hierárquicas tradicionais. De facto, o espaço dos parâmetros fica decomposto em diversas superfícies correspondentes aos diversos sistemas físicos cujos expoentes críticos respectivos dependem do ponto fixo crítico de cada uma das superfícies. É deste modo que se podem classificar em classes de universalidade os diferentes modelos teóricos das transições de fase.

A teoria do Grupo de Renormalização liga-se ao quadro geral da teoria dos sistemas dinâmicos (para uma panorâmica completa – e matematicamente exigente – da teoria dos sistemas dinâmicos *cf.*, por exemplo, Glendinning, 1994). Os seus conceitos fundamentais são os de iteração de uma função e de estado atractor (ponto fixo). É nesse quadro que se pode dar um sentido rigoroso à universalidade. As propriedades de universalidade tornam-se patentes quando se reduz consideravelmente a complexidade interna do sistema (diminuição dos graus de liberdade da rede inicial, no caso da teoria do G. R.). Emergem então certas propriedades macroscópicas de ordem global que se tornam invariantes por relação às dinâmicas internas; invariantes, por exemplo, por relação ao conjunto de relações quânticas "implicitamente" presentes no Hamiltoniano do sistema. Essas propriedades macroscópicas possuem mesmo determinações quantitativas de uma extrema precisão (valor dos expoentes críticos), não podendo deixar de impressionar o facto de esses números puros dependerem apenas duas propriedades geométricas tão gerais quanto d e n. Talvez mesmo mais do que na Física Clássica, existe aqui um sentido profundo da célebre declaração de Galileu segundo "a Natureza está escrita em caracteres geométricos". De um ponto de vista menos filosófico, não é de mais insistir que temos aqui a emergência, a partir da física microscópica, de estados macroscópicos invariantes agindo em escalas temporais maiores. Mas é fundamental compreender-se que, embora seja certo que os estados macroscópicos são em certo sentido causados pelos fenómenos físicos microscópicos subjacentes, não se trata de modo algum de um reducionismo clássico. Os estados macroscópicos são sem dúvida estados físicos, mas são estados físicos invariantes e autónomos por relação aos estados microscópicos. Estes últimos não causam o valor dos expoentes críticos. É apenas a geometria que determina o tipo genérico do sistema crítico (a sua classe de universalidade), bem como o expoente crítico que o caracteriza quantitativamente. Encontramos aqui uma física não directamente ligada ao conceito de energia de uma partícula elementar, e na qual o conceito dominante é o de comportamento crítico macroscópico.

III

A teoria dos fenómenos críticos permite desenvolver em bases "fisicalistas" (mas não reducionistas) um quadro geral acerca da emergência de estados macroscópicos da matéria. É pois normal que ela se tenha tornado um quadro paradigmático para uma vasta classe de fenómenos. É o caso da cognição. A concepção da mente enquanto fenómeno crítico fez-se, inicialmente, através da introdução do conceito de rede neuronal artificial. Refira-se que também certas insuficiências da Inteligência Artificial clássica (como a dificuldade de modelizar, utilizando os formalismos lógicos, as situações contextuais e a aprendizagem) constituíram um forte impulso à investigação em redes neuronais.

Eis um esquema simplificado de uma rede neuronal:



Por analogia com o funcionamento real do cérebro, os nós da rede representam os neuróneos, enquanto as ligações entre eles representam ligações sinápticas. Existe um primeira camada de neuróneos, dita camada de entrada (*inputs*), após o que a rede desenvolve uma dinâmica que vai determinar um certo estado nos neuróneos de saída (*outputs*). Em termos abstractos, uma rede neuronal formal é uma função que associa certos valores do output a certos valores do *input*.

Explicitemos de forma um pouco mais desenvolvida o conceito de rede neuronal.

Uma rede neuronal consiste em N neurónios formais que podem assumir certos estados internos $S \in \{0,1\}$ ou $S \in [0,1]$. Os neurónios encontram-se ligados entre si por certos pesos ou ligações sinápticas W (análogos às ligações sinápticas reais). A associação entre neurónios de entrada e neurónios de saída pode ser definida por:

$$(2)\ s_i = S_j\, W_{ij}\, s_j$$

Temos um neurónio i no estado s_i (activo ou inactivo) que recebe certos pesos ou conexões W provindo de outros neurónios, j, que se encontram nos estados $s_1,...,s_n$. O neurónio s_i muda o seu estado ao computar a soma dos pesos dos neurónios $s_1,...s_n$ (por exemplo, passa a activo se a soma dos pesos é suficientemente elevada). A transição dos estados dos neurónios pode ser definida por uma função do tipo:

$$(3)\ s_i = f(A_i) \qquad \text{com } A_i = S_j\, W_{ij}\, s_j$$

onde f pode ser uma função sigmoidal $1/(1+e^{-x})$ no caso em que $S \in [0.1]$.

Codificada uma certa forma ao nível dos neurónios de entrada (codificação representada por um vector), quer-se que a rede restitua na saída uma forma correspondente à forma dada em entrada. Noutros termos, uma rede deve poder *aprender e classificar formas*. O problema consiste então em, dados os estados internos S, encontrar os pesos adequados W que permitam essa aprendizagem. Isto é, trata-se de garantir que a rede converge efectivamente para a forma desejada. Os algoritmos mais conhecidos que permitem encontrar os W adequados são a Lei de Hebb (se a activação de um neurónio j tende a seguir-se à activação de um neurónio i, então a sua ligação sináptica W tende a reforçar-se):

$$W_{ij} = s_i\, s_j \quad (\alpha \text{ é uma constante de normalização})$$

e a retropropagação, a qual consiste em calcular o afastamento entre a forma obtida e a forma desejada, retropropagando o erro para as camadas anteriores de neurónios (para uma análise detalhada dos algoritmos *cf.* Nadal, 1993). Observe-se que a aprendizagem é controlada pelos pesos W e que estes são um parâmetro externo evoluindo numa dinâmica mais lenta que as dinâmicas rápidas e complexas dos estados internos *s*.

Deve notar-se que, enquanto calculadores e processadores de informação, as redes neuronais operam um cálculo local, distribuído e paralelo. Não existe qualquer unidade central que garanta *a priori* a solução do problema que a rede tem de resolver. Na rede apenas se dão interacções de neurónio vizinho para neurónio vizinho. Uma rede resolve o seu problema quando faz emergir um estado invariante global a partir das múltiplas interacções locais W_{ij}.

O primeiro grande modelo de uma rede neuronal foi introduzido por Hopfield em 1982 (Hopfield, 1982). Isso é particularmente relevante, pois Hopfield baseou explicitamente o seu modelo nos sistemas de *spins* já apresentados. O modelo de Hopfield é mesmo equivalente ao modelo mais simples (o modelo unidimensional) de Ising. De facto, no modelo de Hopfield, assume-se (o que é biologicamente implausível) que se a matriz W (análogo do parâmetro K do sistema de Ising correspondente) dos pesos é simétrica, tem-se uma função de energia dada por:

$$(4)\ E = -\frac{1}{2}\, W_{ij}\, s_i\, s_j$$

Com esta equação, Hopfield mostrou que o seu modelo tinha as propriedades que se encontram no modelo de Ising. Em particular, pode-se introduzir uma temperatura T, análoga à verdadeira temperatura termodinâmica, demonstrando-se que a $T=0$ a energia E atinge um estado estacionário representado por um mínimo (local). Esse mínimo corresponde ao tipo de atractor mais simples, um ponto fixo (exactamente como vimos

suceder no modelo de Ising). Cada um dos pontos fixos que o sistema possa atingir corresponde então a uma forma memorizada. A aplicação de algoritmos como a regra de Hebb ao modelo de Hopfield permite não apenas reencontrar todas as formas memorizadas (correspondendo aos diversos mínimos locais que a energia E pode assumir), mas igualmente reencontrar mínimos absolutos, isto é, obter as formas que nós realmente desejamos que o sistema encontre. Em geral, um teorema demonstrado em 1983 por M. Cohen e S. Grossberg prova que uma classe bastante vasta de sistemas não lineares converge para um estado atractor, convergência assegurada por ser possível definir nesses sistemas uma função de Liapounov (isto é, uma função que decresce na bacia de atracção do atractor A e se anula em A) (*cf.*, por exemplo, Patterson, 1996).

Fazendo variar o parâmetro T no modelo de Hopfield é natural que (tal como nos modelos de Ising equivalentes) exista um diagrama de fases. Assim, já se referiu que a $T \rightarrow 0$ existe um estado atractor correspondendo praticamente a qualquer forma que queremos que a rede armazene em memória. Nesse caso, a rede reconhece de modo praticamente infalível todas as formas armazenadas. Por variação (aumento) de T, ultrapassa-se um limiar crítico onde essas formas deixam de ser estáveis. Para nova variação de T, a rede tende a confundir as formas. Finalmente, para T bastante grande ($T \rightarrow \infty$) temos o estado de equiprobailidade ou máxima entropia (desordem) onde a rede não memoriza nada. Existem pois transições de fase entre memória e confusão total (estado de total desordem), passando pelo "esquecimento" (a rede praticamente não retém qualquer forma (Nadal,1993). Reencontramos aqui as rupturas de simetria resultantes da competição entre ordem e desordem já referidas a propósito dos sistemas de Ising; não é de mais salientar que as transições de fase se dão a valores críticos perfeitamente determinados (cálculos em Nadal, 1993).

Portanto, no caso de uma temperatura baixa e, sobretudo, se se assumir a simetria das conexões W entre os neurónios, é certo que o sistema convergirá para um atractor. Mas se essa última condição for abandonada, deixando assim de lado o estrito paralelismo com a física e procurando modelos biologicamente mais plausíveis (as conexões entre os neurónios reais não são simétricas), a dinâmica de uma rede tipo Hopfield torna-se extremamente complexa e a convergência não está mais assegurada à partida. Encontramos aqui o caso mais geral dos sistemas dinâmicos. O comportamento do sistema vai depender de modo decisivo dos valores do parâmetro W e de outros parâmetros como o nível crítico da activação dos neurónios. Esses parâmetros funcionam como variáveis externas controlando os estados internos S. Prova-se então que a rede sofre uma evolução possuindo um conteúdo universal nos sistemas dinâmicos. Para certos valores das variáveis externas, o sistema converge para um ponto fixo. Se esses valores ultrapassam um certo valor crítico, os atractores tornam-se periódicos. A passagem por um outro valor crítico faz o sistema convergir para um atractor quase-periódico (atractor de período irracional). Finalmente, após atravessar um outro valor crítico, o sistema torna-se caótico: existe sensibilidade às condições iniciais, o sistema sofre constantes bifurcações (infinidade de transições críticas), exibe uma invariância fractal e converge em torno do que se chama um atractor estranho. Encontramos aí um cenário de universalidade: existe uma lei de escalamento constante entre as diferentes bifurcações, a qual nos dá um expoente universal apenas dependente de certas propriedades geométricas muito gerais (um modelo explícito foi desenvolvido por Renals & Rohwer, 1990). A ligação às ideias do Grupo de Renormalização torna-se então evidente, tornando-se possível que os métodos da física dos fenómenos críticos possam migrar completamente para o campo das redes neuronais.

IV

Mesmo sem se poder entrar aqui nos detalhes neuro-químicos acerca da estrutura neuronal do cérebro, é patente que as redes neuronais formais visam constituir modelos cada vez mais aproximados das redes neuronais reais (*cf.* Churchland & Sejnowski, 1992, para uma aproximação do cérebro em termos de redes neuronais formais). Nesse contexto, é importante constatar a quase-identidade entre as redes neuronais formais e a teoria dinâmica dos fenómenos críticos. A hipótese fundamental das ciências cognitivas actuais passa então a formular-se em torno da ideia segundo a qual o cérebro é um sistema dinâmico; a sua modelização apela a teoria matemática dos sistemas dinâmicos; trata-se de um dos aspecto do que T. van Gelder chamou a "hipótese dinâmica acerca da cognição" (van Gelder, 1998) Mais, como se voltará a sublinhar, as redes neuronais são exactamente algoritmos de implementação da teoria dos sistemas dinâmicos. Se esta afirmação é evidentemente verdadeira no que respeita à implementação tecnológica das redes (em máquinas funcionando em paralelo e de forma distribuída), conjectura-se que o mesmo se passa no que respeita às *performances* cognitivas naturais. Segue-se dessa conjectura uma outra hipótese fundamental: *a significação corresponde à presença de um atractor, isto é, ela designa o processo de captura de uma dinâmica cerebral por um atractor.*

Deve com efeito recordar-se que nos modelos neuronais existe uma diferença crucial entre os estados transitórios da rede (os cálculos locais que a rede vai realizando em paralelo), e o estado final assimptótico – o atractor. É apenas ao nível dos atractores que se dá a significação (*cf.* Amit 1989 para o desenvolvimento matemático deste ponto de vista). É uma perspectiva acerca da cognição bastante diferente da veiculada pelo funcionalismo simbólico. Utilizando formalismos mais ou menos aparentados às linguagens da lógica formal, o funcionalismo simbólico não pode reconhecer qualquer distinção entre estados transitórios e estados significativos nem tão pouco, devido à incoerência das linguagens simbólicas, estabelece qualquer distinção entre dinâmicas internas evoluindo a tempo rápido e dinâmicas externas macroscópicas evoluindo a tempo lento.

Torna-se então possível propor um novo tipo de funcionalismo. Ele baseia-se nas noções de universalidade e emergência que encontramos nos sistemas de Ising e nas redes neuronais formais. Trata-se de um funcionalismo na medida em que se respeita uma certa independência em relação a certas dinâmicas internas da matéria. Mas ele visa ao mesmo tempo situar-se ao nível das entidades macroscópicas lentas, invariantes e emergentes a partir das dinâmicas internas. Trata-se de uma perspectiva *bottom-up* que já não parte da lógica e do discreto, mas antes do geométrico e do contínuo. Para esse tipo de funcionalismo, a lógica simbólica não é um dado *a priori* mas algo a conquistar no fim de um processo de progressiva naturalização do sentido.

Esse tipo de funcionalismo começou nos últimos anos a ser assumido de forma mais ou menos consciente por todo um conjunto de autores para quem a perspectiva dinâmica acerca da cognição parece ser a mais correcta (ver a grande recolha de Port & van Gelder, 1995). Ele foi sistematizado por J. Petitot (Petitot, 1992, 1993, 1995) no seguimento do conjunto de ideias propostas por R. Thom – agrupadas por vezes sob a designação de "Teoria das Catástrofes". Na terminologia de Petitot, trata-se de um funcionalismo dinâmico visando anular a cisão histórica entre Sentido e Natureza. É importante realçar que esse funcionalismo não deve ser identificado com as ideias filosófico-tecnológicas ligadas às redes neuronais e teorizadas por P. Smolensky sob o nome de conexionismo ou paradigma sub-simbólico (Smolensky, 1988). Enquanto teoria de uma tecnologia, o conexionismo não faz necessariamente parte de um funcionalismo dinâmico da mente. Mais do que procurar de imediato explicar as estruturas lógicas em termos de redes neuronais artificiais, o funcionalismo dinâmico nunca perde de vista a ligação à física, e em particular à teoria dos sistemas dinâmicos. Os conceitos de atractor e de universalidade física fornecem o seu quadro teórico essencial.

Neste artigo panorâmico não podemos entrar nos múltiplos subprogramas de investigação nos quais o funcionalismo dinâmico se ramifica. Um exemplo permitirá contudo perceber a sua estratégia geral. É sabido que a modelização do cérebro através de sistemas dinâmicos enfrenta o problema de o número de neurónios e estados neuronais ser gigantesco. (Aliás, o mesmo é válido no caso das redes neuronais formais.) A modelização tem de fazer intervir um espaço com um número enorme de dimensões. A consequência disso reside no facto de os modelos rapidamente apresentarem trajectórias convergindo para atractores estranhos. De facto, com a dimensão ≥3 podem surgir fenómenos tipo "caos"[5]. Assim, a equação significação = atractor implica a equação significação = atractor estranho. Um funcionalismo dinâmico deve então procurar reduzir essa complexidade e encontrar os invariantes do nível macroscópico + 1 relativamente aos invariantes associados ao nível macroscópico dos atractores estranhos, invariante esse agindo sempre num tempo mais lento que o nível cuja complexidade ele reduz. O ponto é de grande delicadeza matemática. Uma ideia avançada por R. Thom consiste em não considerar os atractores mas sim uma "equivalência de atractores, equivalência definida pela pertença a uma variedade de nível de uma função de Liapounov" (Thom 1990, p. 521). Isto é, trata-se de apenas reter das bifurcações (infinitas) de atractores a sua função de Liapounov, pelo que as trajectórias presentes na região de atracção convergem para o atractor tal como no caso de um atractor simples. A estrutura interna do atractor é pois reduzida, ficando apenas o atractor definido pela existência de uma função de Liapounov. Ter-se-ia assim um invariante por relação à dinâmica complexa dos atractores gerais (caóticos). A estratégia do funcionalismo dinâmico é uma estratégia de redução sucessiva da complexidade das dinâmicas internas. Quanto mais "alto" for o nível de realidade considerado menor é o número de dimensões do espaço que o descreve e maior a escala de acção do tempo. Foi precisamente isso que se viu a propósito do Grupo de Renormalização: redução das dimensões do sistema pela operação de "tome-se uma média do comportamento do sistema original e reescale-se". É o que igualmente encontramos a propósito da modelização do cérebro: reduzam-se as dimensões do sistema e procurem-se "médias" dos atractores estranhos. Deve notar-se que este reducionismo segue uma estratégia inversa da do reducionismo clássico, o qual visa reduzir tudo às dinâmicas internas extremamente complexas.

Qual é o nível a que se situa o funcionalismo dinâmico? O funcionalismo simbólico situa-se ao nível dos invariantes de todos os invariantes, que seria o nível da lógica formal. Ou antes, ele situa-se próximo desse nível, pois são introduzidas linguagens formais sem a transparência de linguagens lógicas como o cálculo de predicados à primeira ordem. Como já se referiu, no quadro de um funcionalismo dinâmico esse invariante lógico é algo mais a conquistar do que um dado *a priori*. O funcionalismo dinâmico é a hipótese segundo a qual diversos tipos de invariantes, cada vez mais gerais e mais "macroscópicos" se podem ir sucedendo uns aos outros. O nível do funcionalismo dinâmico é portanto relativo. Ele passará pelo nível associado ao Grupo de Renormalização aplicado aos fenómenos críticos, sendo aí que a cognição começará a emergir. Mas como esse nível é ainda bastante complexo, deverá operar-se uma nova redução e passar ao nível dos objectos morfologicamente estruturados, nível descrito pela chamada Teoria das Catástrofes Elementares (fora de questão entrar aqui em detalhes matemáticos; já existe um bom e (relativamente) elementar tratamento no manual da autoria de Castrigiano & Hayes, 1993). Esse nível é um invariante por relação às bifurcações gerais dos sistemas dinâmicos (*cf.* Petitot,1992, para detalhes matemáticos e filosofia geral). Em suma, e de um ponto de vista cognitivo, o nível funcional será o da teoria dos sistemas dinâmicos gerais mais a sua teoria fenomenológica, a Teoria das Catástrofes Elementares.

[5] Mais exactamente, trata-se de uma consequência da natureza oscilatória do cérebro, provando-se matematicamente que um regime periódico com três frequências é instável, estabilizando-se num regime caótico. Walter Freeman apresentou bastante evidência experimental nesse sentido. *Cf.*, por exemplo, o seu 1993.

Tal como no caso do funcionalismo clássico, o problema da implementação torna-se então central para um funcionalismo dinâmico. Viu-se que o funcionalismo clássico resolvia metade desse problema através da analogia *software-hardware*, restando obscuro o problema da implementação das linguagens simbólicas em cérebros reais. Ora, se o nível do funcionalismo dinâmico é dado pela teoria geral dos sistemas dinâmicos, a questão é a de saber como se implementa essa teoria em cérebros. Noutros termos, quais são os algoritmos correspondentes ao nível funcional dinâmico? Podem ser vários, mas seguramente que uma das classes principais desses algoritmos é precisamente fornecida pelas redes neurais formais[6]. De facto, como se indicou, uma rede neuronal formal é precisamente um algoritmo implementando (em cérebros ou em certo tipo de máquinas) sistemas dinâmicos.

É exactamente por possuir algoritmos correlativos que o funcionalismo dinâmico é um monismo, contraposto ao dualismo típico do funcionalismo simbólico. Este ponto tem interesse filosófico, sendo com ele que terminamos.

V

O ponto que desejamos sublinhar pode ser visto a partir de um aspecto crucial do célebre debate entre J. Fodor e P. Smolensky acerca do estatuto das linguagens simbólicas por relação às estruturas sub-simbólicas ou conexionistas. A crítica fundamental de Smolensky ao funcionalismo simbólico de Fodor consiste em sustentar que "as regras simbólicas são reais no sentido de governarem a semântica e a função computada, mas não são reais no sentido de participarem na história causal, capturável por um algoritmo, do mecanismo interno que computa essas funções" (Smolensky, 1995, p. 225). Utilizando distinções chomskianas, as estruturas ideais simbólicas seriam, segundo Smolensky, teorias da competência e não da *performance*. Elas seriam teorias da capacidade simbólica humana, mas não possuiriam poder causal por não estarem realmente implementadas. As performances cognitivas não seriam simbólicas, mas sim conexionistas. Não se recusa a realidade (no sentido de realidade ideal) às estruturas simbólicas, mas afirma-se que a causalidade dos actos mentais – o nível algorítmico – é subsimbólico. Resulta daí que os algoritmos das dinâmicas mentais não são algoritmos tipo LISP mas sim algoritmos neuronais do tipo dos acima apresentados. No seu debate com Fodor, Smolensky está claramente ansioso por recuperar no quadro dos algoritmos sub-simbólicos ou neuronais as descrições simbólicas de alto nível, o que o conduz à ideia (compatível com um funcionalismo dinâmico) segundo a qual as estruturas simbólicas seriam invariantes ideais das *performances reais* representadas pelos algoritmos conexionistas.

Seguindo a mesma linha de raciocínio, mas evitando a questão (delicada, se bem que decisiva) da emergência do simbólico, resulta do que até agora foi exposto que os algoritmos conexionistas implementam, ao nível das *performances cognitivas*, não as estruturas simbólicas mas sim a teoria geral dos sistemas dinâmicos.[7] Nessa medida, os algoritmos conexionistas seriam instanciações (no sentido da distinção *type/token*) mentais de certas estruturas ideais. Essas estruturas não seriam constituídas por linguagens simbólicas mas sim pela teoria geral dos sistemas dinâmicos. Falamos aqui dessa *teoria*, necessariamente um produto cognitivo como qualquer outra actividade mental. Enquanto teoria ou modelo, a teoria dos sistemas dinâmicos é um tipo invariante que se instancia instanciada cognitivamente nos actos mentais através dos algoritmos conexionistas.

As anteriores afirmações implicam um monismo que em certo sentido retoma a velha problemática da analogia microcosmos-macrocosmos. A analogia não se baseia apenas no facto de a teoria dos sistemas dinâmicos aplicada aos fenómenos críticos estabelecer tecnicamente um isomorfismo entre cérebro e fenómenos gerais de transição de fases, ao que se poderia acrescentar que existe aí um caso de universalidade. A hipótese que se avançou consiste em antes dizer que a teoria acerca dos processos de emergência na natureza é ela própria uma teoria que possui realidade cognitiva; possui-a não enquanto teoria, não enquanto modelo de fenómenos naturais, mas enquanto algo que possui eficácia na mente através de um algoritmo. Noutros termos, as teorias que servem para representar o mundo estariam elas próprias mental e neuronalmente implementadas: elas emergiriam, enquanto teorias ideais e objectivas, a partir dessa sua realidade cognitiva. A Natureza simular-se-ia na mente, num processo de mimesis recíproco. A circularidade não existe aqui (ao contrário do que sempre sucedeu com as teorias da analogia microcosmos/macrocosmos) na medida em que se distingue claramente entre a teoria e a realidade cognitiva que ela possui ao nível das dinâmicas internas da mente. Enquanto ideal, a teoria emerge a partir das dinâmicas mentais e é um seu invariante; enquanto real (enquanto instância), é um processo cognitivo análogo a qualquer outro processo natural.

O funcionalismo dinâmico envolve pois a tese monista (um monismo naturalista, Petitot, 1993, p. 79) e um realismo. Distingue-se pois dos dualismos e solipsismos simbólicos. Ele é no fundo uma tentativa de precisar algumas das ideias do criador de muitas das concepções apresentadas ao longo do artigo, R. Thom. Em 1968, este propunha

> *criar uma teoria da significação cuja natureza seja tal que o próprio acto de conhecer seja uma consequência da teoria.* (Thom, 1980, p. 179)

No quadro de um funcionalismo dinâmico, seriam as ciências cognitivas contemporâneas que permitiriam criar uma tal teoria da significação.

[6] Uma discussão mais aprofundada deste ponto teria de incidir acerca da maior ou menor plausibilidade biológica de alguns dos algoritmos das redes neuronais: se algoritmos como a lei de Hebb possuem essa plausibilidade, já o mesmo não sucede com o algoritmo da retropropagação.

[7] A exposição aqui feita pode deixar obscurecido o facto de existirem diversos tipos de redes neuronais, alguns deles tendo mais pontos de contacto com os algoritmos simbólicos do que propriamente com os sistemas dinâmicos. A argumentação desenvolvida no texto tem subjacente essencialmente as redes neuronais recorrentes e, mais em geral, as redes neuronais onde a ideia de atractor é nuclear.

Referências bibliográficas

AMIT, D.: *Modeling Brain Function*, Cambridge University Press, Cambridge, 1989.

BLOCK N.: "The Computer model of the Mind", in A. Goldman (ed.), *Readings* in the Philosophy and Cognitive Science, Mit Press, Cambridge, Mass, 1993.

CASTRIGIANO, D. & HAYES, S.: *Catastrophe Theory*, Addison-Wesley, Reading, 1993.

CHURCHLAND P. & SEJNOWSKI,T.: *The Computacional Brain*, MIT Press, Cambridge, 1992.

DUPUY, J. P.: *Aux Origines des Sciences Cognitives*, Ed. de la découverte, Paris, 1994.

FISHER. M.: *Scaling, Universality and Renormalization Group Theory*, Springer-Verlag, Berlin, 1983.

FODOR, J.: *The Language of Thought*, Harvard U. P., Cambridge, Mass, 1975.

FODOR, J.: *Representations: Philosophical Essays on the Foundations of Cognitive Science*, MIT Press, Cambridge, Mass, 1983.

FREEMAN, W.: "Chaos in the CNS: theory and practice", in Ventriglia, F., (ed.) *Neural Modeling and Neural Networks*, Pergamon Press, New York, 1993, pp. 185-216.

GLENDINNING, P.: *Stability, Instability and Chaos – an Introduction to the theory of nonlinear differential equations*, Cambridge University Press, Cambridge, 1994.

HODGES, A.: *Alan Turing: The Enigma*, Burnett, Londres, 1983.

HOPFIELD, J.: "Neural Networks and Physical Systems with Emergent Collective Computational Abilities", *Proceedings of the National Academy of Sciences*, USA, 79: 2554-2558, 1982

KLENNE, S.: *Logique Mathématique*, J. Gabay, Paris, 1987.

LOEWER, B., & REY, G.: (eds), *Meaning in Mind – Fodor and his Critics*, Blackwell, Oxford, 1991.

LOPES, M.: *Fenómenos Críticos – da Estática à Dinâmica*, tese da Faculdade de Ciências de Lisboa, 1988.

PATTEE, H. H: "Evolving self-reference: matter, symbols, and semantic closure" in *Communication and Cognition – Artificial Intelligence*, 1996, Vol. 12, Nos. 1-2, 9-27.

PATTERSON, D.: *Artificial Neural Networks*, Prentice Hall, Singapura, 1996.

PENROSE, R.: *The Emperor's New Mind*, Oxford University Press, Oxford, 1989.

PETITOT, J.: *Physique du Sens*, C.N.R.S., Paris, 1992.

PETITOT, J.: "Phénoménologie naturalisée et morphodynamique : la function cognitive du synthétique a priori", *Intellectica*, 17, 1993, 79-126.

PETITOT, J.: Morphodynamics and Atrractor Syntax: Constituency in Visual Perception and Cognitive Grammar, in Porter & van Gelder, 1995.

PORT, R., & VAN GELDER, T.: *Mind as Motion – Exploration in the Dynamics of Cognition*, MIT Press, Cambridge, Mass, 1995.

PUTNAM, H.: "Minds and Machines", in H. Putnam, *Philosophical Papers*, Vol. II, Cambridge University Press, Cambridge, 1975, pp. 362-385.

PYLYSHYN, Z.: *Computation and Cognition. Toward a Foundation for Cognitive Science*, MIT Press, Cambridge, Mass, 1984.

RENALS S., & ROHWER R.: "A Study of Network Dynamics", *Journal of Statistical Physics*, 58, 1990, 5/6, 825-848.

ROSA, A. M: *Le Concept de Continuité chez C.S. Peirce*, E.H.E.S.S., Paris, 1993.

ROSA, A. M.: *Ciência, Tecnologia e Ideologia Social*, U.L.H.T., Lisboa, 1995.

SEARLE J.: *The Rediscovery of the Mind*, MIT Press, Cambridge, Mass, 1992.

SMOLENSKY, P.: "On the Proper Treatment of Connectionism", *The Behavioral and Brain Sciences*, 11, 1988, 1-23.

SMOLENSKY, P.: "Reply: Constituent Structure and Explanation in a Integrated Connectionist/Symbolic Cognitive Architecture" *in* MacDonald, C., & MacDonald, C. (eds.). *Connectionism – Debates on Psychological Explanation*, Blackwell, Oxford, 1995.

THOM, R.: *Modèles mathématiques de la Morphogenèse*, C. Bourgois, Paris, 1980.

THOM, R.: *Apologie du Logos*, Hachette, Paris, 1990.

VAN GELDER, T. J.: The dynamical hypothesis in cognitive science, a aparecer in *Behavioral and Brain Sciences*, 1998.