

Breve História do Conexionismo entre o Fim do Séc. XIX e ~1950

O objetivo desta aula é *preparar* o aluno para o resto do curso. Os conceitos a serem apresentados nesta disciplina são novos para a maioria dos alunos e muitos têm dificuldade em apreender o conteúdo se ele for transmitido com certa pressa. Portanto, a melhor maneira de seguir adiante é ir devagar.

Outra boa maneira de apresentar uma matéria nova é contar um pouco da sua história. A história de qualquer ramo da ciência é muito complexa; há vários atores participando – em diferentes partes do globo – e as maneiras como eles interagem e divulgam suas idéias e seus trabalhos, assim como as razões pelas quais algumas dessas idéias e alguns resultados dos seus trabalhos são aceitos e outros são recusados são difíceis de ser entendidas de uma forma puramente objetiva e linear. Como o propósito deste curso não é destrinchar a história do conexionismo, vamos nos ater apenas a alguns personagens principais e seguir a linha histórica “padrão” adotada pela maioria dos livros sobre conexionismo e redes neurais. Uma boa referência para esta aula é o livro editado por Anderson e Rosenfeld (*Neurocomputing*, vol. 1), citado nas referências da aula 1.

Outro motivo para esta aula estar sendo dada agora é que ela vai se concentrar nos elementos *essenciais* do conexionismo, deixando muito da biologia usada para justificá-lo de fora. Lembre-se que foi dito na aula 1 que o apelo do conexionismo para os psicólogos e cientistas cognitivos reside mais na maneira como ele aborda a computação do que na sua plausibilidade biológica. Sendo assim, considero mais conveniente apresentar para vocês alguns conceitos fundamentais do conexionismo como eles foram concebidos antes da década de 1960, a partir da qual a neurobiologia tem passado por uma dramática evolução¹.

¹ Segundo Squire *et al.* (2008), o termo “neurociência” apareceu na metade dos anos 1960 como uma sinalização do início de uma era em que as diferentes disciplinas devotadas ao estudo do cérebro (neuroanatomia, neurofisiologia, neuroquímica, psicologia, etc) estariam trabalhando juntas de forma cooperativa, compartilhando uma linguagem comum, conceitos comuns e objetivos comuns para entender a estrutura e a função do cérebro normal e anormal. A neurociência é, no momento, uma das áreas científicas de mais rápido crescimento no mundo. A primeira reunião anual da *Society for Neuroscience*, em 1971, contou com a participação de 1100 pessoas. A 40ª. reunião anual da mesma sociedade em 2010 teve 31.975 participantes.

William James

O filósofo e psicólogo estadunidense William James (1842-1910) foi um dos pioneiros da psicologia científica moderna. A sua formação básica foi em medicina e fisiologia, o que lhe deu condições técnicas para investigar os fenômenos mentais e tentar entendê-los em termos do seu substrato neurobiológico. Dos vários livros escritos por James, o principal foi *The Principles of Psychology* (1890), que pode ser lido na internet em <http://psychclassics.yorku.ca/James/Principles/>².

É interessante reproduzir aqui alguns dos comentários feitos por Anderson e Rosenfeld (op. cit, capítulo 1, pp. 1-2) sobre as idéias e textos de James³:

“Como James enfatiza em várias partes [do seu livro], o cérebro não está construído para pensar abstratamente – ele está construído para garantir sobrevivência no mundo. Ele possui muitas características de uma boa solução de engenharia aplicada à operação mental: faça um trabalho tão bom quanto puder, sem gastar muito e com o que puder obter facilmente. Se isso significar usar soluções ad hoc, com menos generalidade que se poderia gostar, então que seja assim. Vivemos em um mundo particular, e somos construídos para trabalhar nele e com ele. Como James escreveu”, (o trecho a seguir é do próprio James, da versão resumida do *The Principles of Psychology*; os itálicos são de Anderson e Rosenfeld)

“Os fatos mentais não podem ser corretamente estudados separados do ambiente físico do qual eles tomam conhecimento. A grande falha da velha psicologia racional foi estabelecer a alma como um ser espiritual absoluto com certas faculdades intrínsecas pelas quais as várias atividades de lembrança, imaginação, raciocínio, vontade, etc. eram explicadas, quase sem referência às peculiaridades do mundo com o qual essas atividades lidam. Mas o insight mais rico dos dias modernos percebe que nossas faculdades interiores são

² Para mais sítios na internet onde obras de William James podem ser lidas, ver a página sobre ele na Wikipedia: http://pt.wikipedia.org/wiki/William_James.

³ A tradução desse texto, assim como as de todos os outros apresentados neste curso, foi feita livremente pelo professor.

5945851-1 – Psicologia Conexcionista – Antonio Roque – Aula 2
antecipadamente *adaptadas* às características do mundo no qual habitamos. ... A mente e o mundo, em resumo, evoluíram juntos, e em consequência são algo como um ajuste mútuo”.

Anderson e Rosenfeld continuam mais adiante (op. cit. p. 2):

“Uma importante implicação disso é que não devemos nos surpreender se os cérebros não puderem computar tudo, mas apenas um subconjunto pequeno, mas útil, de problemas. Os cérebros são apenas tão poderosos quanto eles têm que ser, e são comumente surpreendentemente específicos. O tipo de organização cerebral que nós parecemos ter é muito pobre para fazer operações aritméticas e da lógica formal; quando o cérebro estava evoluindo, houve oportunidades limitadas para fazer as duas coisas. Mas a habilidade para formar conceitos, para perceber que coisas diferentes eram exemplos da mesma coisa, era realmente importante, assim como eram as habilidades para formar associações um tanto arbitrárias, para fazer boas suposições, e assim por diante. É surpreendente que nossos cérebros, que evoluíram para propósitos como esses, sejam também capazes de fazer operações aritméticas e de lógica formal”.

“A coisa mais interessante do capítulo sobre associação (do livro *The Principles of Psychology*) é que ele apresenta um modelo mecanístico detalhado de associação que é quase idêntico em estrutura a modelos posteriores de redes neurais associativas” (que só apareceriam por volta da década de 1970).

O modelo mecanístico de James para associação se baseia no seguinte princípio, postulado por ele como a lei elementar da associação (*The Principles of Psychology*, p. 566):

Quando dois processos cerebrais elementares tiverem estado ativos juntos ou em sucessão imediata, um deles, ao re-ocorrer, tende a propagar sua excitação para o outro.

Este princípio, no entanto, não é suficiente para a construção de um modelo mecanístico. É necessário detalhar melhor como deve ser a sua operação. Nas palavras do próprio James (op. cit., pp. 452-454; a ênfase em itálico no segundo parágrafo é do próprio James):

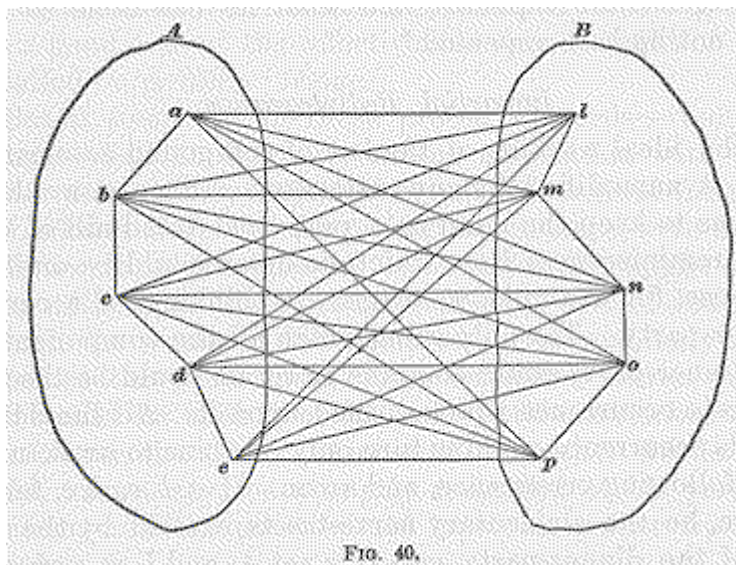
“Mas, na realidade, todo processo elementar já se encontrou em tempos diferentes excitado em conjunção com *muitos* outros processos, e isso por causas externas inevitáveis. Qual desses outros ele deve despertar agora se torna um problema. Será *b* ou *c* o próximo a ser estimulado pela presença de *a*? Devemos fazer um postulado a mais, baseado no fato da *tensão* no tecido nervoso, e no fato da somação de excitações, cada uma incompleta ou latente em si mesma, sobre uma resultante aberta. O processo *b*, ao invés de *c*, despertará, se em adição ao trato vibrante *a* algum outro trato *d* estiver em estado de sub-excitação, e anteriormente esteve excitado com *b* sozinho e não com *a*. Em resumo, podemos dizer”:

“A quantidade de atividade em algum ponto no córtex do cérebro é a soma das tendências de todos outros pontos a descarregar sobre ele, tais tendências sendo proporcionais (1) ao número de vezes que a excitação de cada outro possa ter acompanhado a do ponto em questão; (2) à intensidade de tais excitações; e (3) à ausência de qualquer ponto rival funcionalmente desconectado do primeiro ponto, sobre o qual a descarga possa ser desviada”.

A concepção de James sobre associação entre padrões cerebrais é incrivelmente moderna, muito similar à do conexionismo dos dias atuais. A sua lei elementar da associação é uma antecipação da chamada “lei de Hebb” (que veremos mais adiante), que é uma das principais regras de aprendizado do conexionismo⁴.

⁴ Na realidade, Aristóteles (384 a.C. – 322 a.C.) já tinha uma teoria para as associações mentais baseada em quatro “leis” de aprendizado (lei da similaridade, lei do contraste, lei da frequência e lei da contigüidade). O conceito de “associação” como um dos mecanismos centrais para a formação de pensamentos foi muito considerado pelos filósofos empiristas ingleses John Locke (1632 – 1704), George Berkeley (1685 – 1753), David Hartley (1705 – 1757) e David Hume (1711 – 1776). Hartley, em particular, chegou a propor um modelo físico para a formação de associações segundo o qual os estímulos que atingem os nossos órgãos sensoriais criam pequenas vibrações que se propagam pelos nervos chegando ao cérebro. Essas vibrações ficariam associadas com quaisquer sentimentos ou idéias que o cérebro estiver “pensando” no momento da sua chegada, de maneira que futuras ocorrências dos mesmos estímulos provocariam as mesmas idéias e pensamentos.

Uma figura do livro de James (Figura 57, reproduzida abaixo) mostra como ele imaginava a associação entre dois pensamentos distintos, *A* e *B*. Observe que cada pensamento é formado por sub-unidades (representadas por letras minúsculas, *a*, *b*, *c*, etc) e que a associação entre os pensamentos é feita por conexões entre essas sub-unidades. Esta figura é muito similar aos diagramas modernos usados para representar redes neurais: os “estados da rede” são representados por padrões distribuídos fortemente interconectados.



O Modelo de McCulloch e Pitts

O primeiro modelo de redes neurais artificiais foi proposto por Warren S. McCulloch (1898 – 1969) e Walter Pitts (1923 – 1969), em um artigo publicado em 1943: “A logical calculus of the ideas immanent in nervous activity”, *Bulletin of Mathematical Biophysics*, 5: 115-133.

McCulloch era psiquiatra e neuroanatomista e Pitts era lógico e matemático. Ambos faziam parte de um dos primeiros grupos do mundo dedicado ao estudo da biologia teórica, criado pelo físico russo-estadunidense Nicolas Rashevsky (1899 – 1972) na Universidade de Chicago.

Em 1943, o conhecimento sobre os neurônios biológicos era muito limitado. As bases iônicas e elétricas da atividade neuronal eram ainda incertas, porém já se sabia da existência de potenciais de ação e da sua natureza “tudo-ou-nada”.

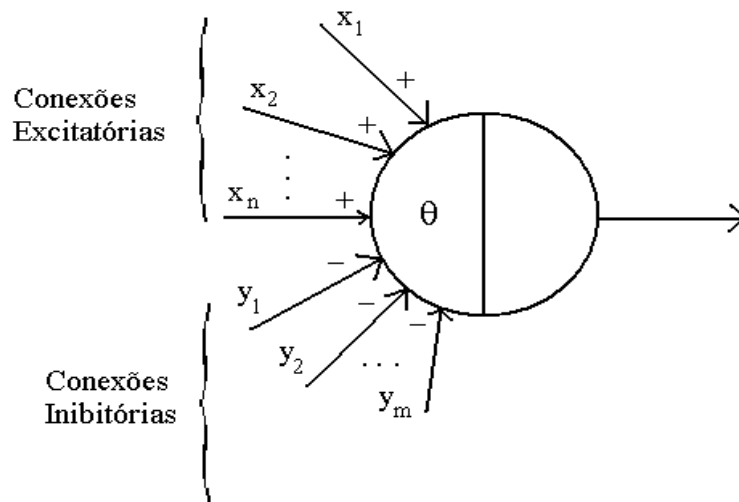
A propriedade “tudo-ou-nada” da resposta de um neurônio sugeriu a McCulloch e Pitts que um neurônio pode representar uma proposição lógica formal (que, necessariamente, só pode ser de dois tipos: verdadeira ou falsa). Desta maneira, as relações existentes entre as atividades de neurônios interconectados em uma rede podem representar relações entre proposições lógicas.

McCulloch e Pitts propuseram um modelo de sistema neural em que as unidades básicas, os neurônios, são bastante simples no seu funcionamento. A riqueza de processamento e a capacidade computacional do sistema são decorrentes da *conectividade* entre esses elementos simples. Pode-se dizer que o modelo de McCulloch e Pitts foi o primeiro modelo conexionista formulado matematicamente.

O Modelo de McCulloch e Pitts está baseado em cinco hipóteses:

1. A atividade de um neurônio é binária, ou seja, a cada instante o neurônio, ou está disparando (atividade 1), ou não está disparando (atividade 0);
2. A rede neural é constituída por linhas direcionadas, sem pesos, ligando os neurônios. Essas linhas (inspiradas nas sinapses) podem ser excitatórias ou inibitórias;
3. Cada neurônio tem um limiar fixo θ , de maneira que ele só dispara se a entrada total chegando a ele, num dado instante, for maior ou igual a θ ;
4. A chegada de uma única sinapse inibitória num dado instante evita absolutamente o disparo do neurônio, independentemente do número de sinapses excitatórias que estejam chegando conjuntamente com a sinapse inibitória;
5. Um sinal leva uma unidade de tempo para passar de um neurônio da rede para outro. Isso procura reproduzir o atraso sináptico.

A figura abaixo mostra o modelo de neurônio de McCulloch-Pitts.



Se, num dado instante de tempo, pelo menos um dos y_i s for 1, o neurônio é inibido. Caso nenhum dos y_i s seja igual a 1, o neurônio calcula a soma $\sum x_i$ e a compara com θ . Se $\sum x_i \geq \theta$, ele dispara; caso contrário, ele não dispara.

Em seu artigo de 1943, McCulloch e Pitts provaram o seguinte teorema:

“Qualquer expressão lógica finita pode ser implementada por uma rede de neurônios do tipo definido pelas cinco hipóteses acima”.

A prova deste teorema foi obtida com o uso de regras da lógica formal, o que torna o artigo de McCulloch e Pitts bastante árduo de ser lido. Porém o impacto deste resultado foi muito grande na época. Em outras palavras, o que o teorema diz é que:

“Qualquer procedimento, ou *computação*, pode ser implementado por uma rede de neurônios de McCulloch e Pitts”.

Este resultado mostrou que uma rede feita de unidades bastante simples pode ter um poder computacional enorme.

Como as unidades do modelo são baseadas nos neurônios do cérebro, isto sugere que o próprio cérebro tem o potencial de executar computações lógicas complexas. Isto nós todos sabemos, mas o artigo de McCulloch e Pitts foi o primeiro a mostrar que pode ser possível entender matematicamente as computações feitas pelo cérebro.

O modelo de McCulloch e Pitts não teve muita influência na neurobiologia. As evidências experimentais que começavam a se acumular a partir da década de 1940 indicavam que o neurônio é mais complexo do que um simples elemento binário com um limiar e isso fez com que os neurobiólogos não dessem muita atenção ao trabalho de McCulloch e Pitts.

O grande impacto do modelo de McCulloch e Pitts deu-se na ciência da computação. O pai dos modernos computadores digitais, o matemático húngaro-estadunidense John von Neumann (1903 – 1957), foi bastante influenciado pelo trabalho de McCulloch e Pitts, pois ele percebeu o grande poder computacional que um sistema composto por unidades lógicas simples possui.

Em 1945, von Neumann escreveu o seguinte em um relatório ao exército norte-americano sobre o EDVAC (*Electronic Discrete Variable Automatic Computer*), um dos primeiros computadores digitais:

Vale a pena mencionar que os neurônios dos animais superiores são elementos (de tipo relé) no sentido acima. Eles são de característica tudo-ou-nada, isto é, têm dois estados: quiescente e excitado. ... Um neurônio excitado emite o estímulo padrão ao longo de muitas linhas (axônios). Uma dessas linhas pode, entretanto, se conectar de duas maneiras diferentes ao próximo neurônio: Primeiro, em uma sinapse excitatória, de tal maneira que o estímulo causa a excitação daquele neurônio. Segundo, em uma sinapse inibitória, de tal maneira que o estímulo evita absolutamente a excitação daquele neurônio por qualquer estímulo em qualquer outra sinapse excitatória. O neurônio também tem um tempo de reação definido, entre a recepção de um estímulo e a emissão do estímulo causado por ele, o atraso sináptico.

Seguindo W. Pitts e W. S. McCulloch ... ignoramos os aspectos mais complicados do funcionamento de um neurônio: limiares, somação temporal, inibição relativa, ... etc ... Pode-se ver claramente que este neurônio simplificado pode ser imitado por relés telegráficos ou por válvulas.

Este texto deixa bem claro que von Neumann fazia uma analogia direta entre uma rede de neurônios de McCulloch e Pitts e a arquitetura e os princípios de funcionamento de um computador digital.

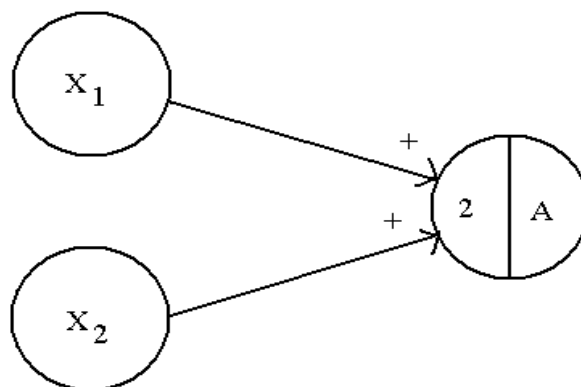
Vejamos a seguir alguns exemplos de aplicações do modelo de McCulloch e Pitts.

Exemplos de funções lógicas booleanas de duas variáveis implementadas por redes de neurônios de McCulloch e Pitts:

1. Função lógica “e”: A tabela de verdade desta função está dada abaixo

X_1	0	0	1	1
X_2	0	1	0	1
A	0	0	0	1

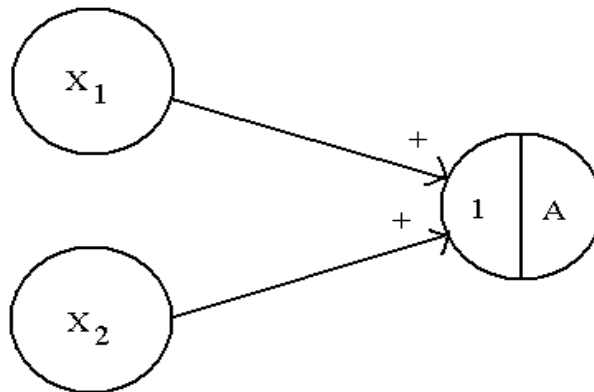
$A = 1$ se X_1 e X_2 forem iguais a 1.



2. Função lógica “ou”: A tabela de verdade desta função está dada abaixo

X_1	0	0	1	1
X_2	0	1	0	1
A	0	1	1	1

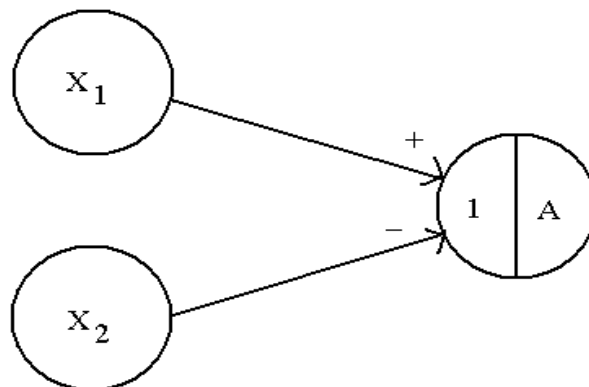
$A = 1$ se X_1 ou X_2 , ou ambos, forem iguais a 1.



3. Função lógica “e não”: A tabela de verdade desta função está dada abaixo

X_1	0	0	1	1
X_2	0	1	0	1
A	0	0	1	0

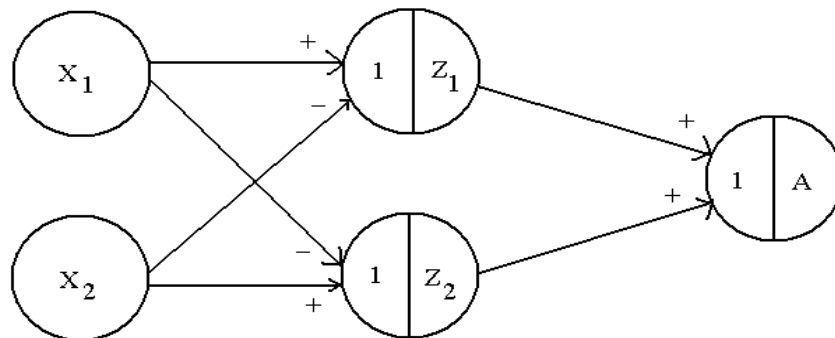
$A = 1$ somente se $X_1 = 1$ e $X_2 = 0$.



4. Função lógica “xou”, dita “ou exclusivo: A tabela de verdade desta função está dada abaixo

X_1	0	0	1	1
X_2	0	1	0	1
A	0	1	1	0

$A = 1$ se $X_1 = 1$ ou $X_2 = 1$, mas não se ambos forem iguais a 1.

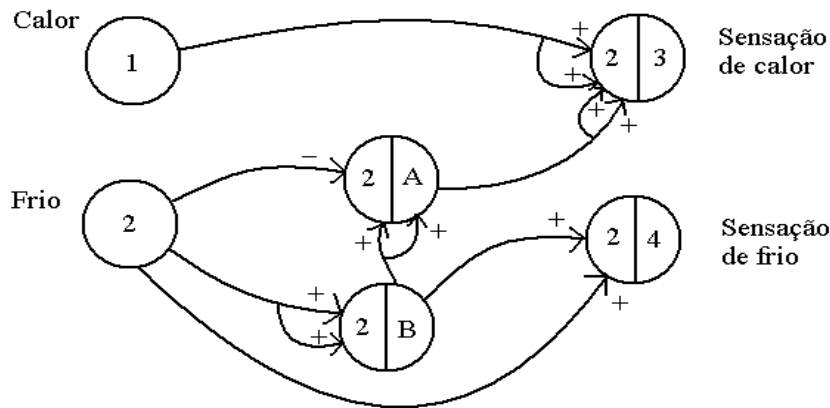


A função lógica XOU pode ser expressa em termos das funções lógicas “E NÃO” e “OU” como: $X_1 \text{ XOU } X_2 \leftrightarrow (X_1 \text{ E NÃO } X_2) \text{ OU } (X_2 \text{ E NÃO } X_1)$. Note que a primeira parte da rede acima consiste de duas redes do tipo “E NÃO” e a segunda parte consiste de uma rede do tipo “OU”.

Exemplo de aplicação do modelo de McCulloch e Pitts à modelagem do fenômeno de percepção de calor e de frio pela pele, apresentado por eles mesmos em 1943:

Quando um objeto frio é encostado na pele e logo retirado, a pessoa tem a sensação de calor (de queimadura). Já quando um objeto frio é encostado na pele por um longo tempo, a pessoa tem a sensação de frio.

Seja a rede da figura abaixo: o neurônio 1 é um receptor de calor da pele, o neurônio 2 é um receptor de frio, o neurônio 3 causa sensação de calor quando ativado e o neurônio 4 causa sensação de frio quando ativado. Os neurônios A e B são interneurônios.



A tabela abaixo mostra a seqüência de eventos em um caso em que no instante inicial, $t = 1$, um objeto frio é encostado à pele e retirado logo em seguida. Isso provoca uma sensação de calor em $t = 4$.

Tempo	1	2	A	B	3	4
1	Não	Sim	Não	Não	Não	Não
2	Não	Não	Não	Sim	Não	Não
3	Não	Não	Sim	Não	Não	Não
4	Não	Não	Não	Não	Sim	Não

Já a tabela abaixo mostra a seqüência de eventos em um caso em que o objeto frio fica em contato com a pele continuamente, provocando, em $t = 3$, uma sensação de frio.

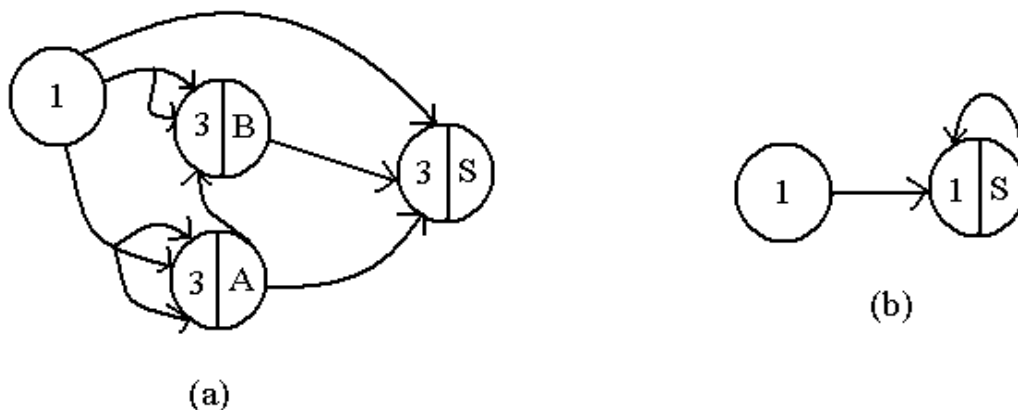
Tempo	1	2	A	B	3	4
1	Não	Sim	Não	Não	Não	Não
2	Não	Sim	Não	Sim	Não	Não
3	Não	Sim	Não	Sim	Não	Sim

Este modelo de McCulloch e Pitts, embora bastante simples, já inclui alguns elementos importantes dos modernos modelos de redes neurais.

Por exemplo, ele contém três tipos de unidades: neurônios de entrada, neurônios de saída e neurônios “ocultos”. As unidades de entrada recebem os sinais, ou estímulos, vindos do ambiente externo à rede neural. As unidades de saída fornecem a resposta da rede ao padrão particular de estímulos que chega às unidades de entrada. As unidades ocultas (este termo foi introduzido bem mais tarde, em 1986, por Rumelhart e McClelland) não têm contato direto com o ambiente externo à rede (nem recebem, nem enviam sinais para ele), só interagindo com outros neurônios da própria rede.

Pode-se pensar na tríade, neurônios de entrada, de saída e ocultos, como inspirada na situação do cérebro, constituído por neurônios receptores, motores e interneurônios. Porém, nem sempre o que se espera dos neurônios de saída é uma resposta motora. Ela pode ser um padrão de atividade que represente algum estado cognitivo, como uma categorização ou uma emoção, por exemplo.

McCulloch e Pitts também se preocuparam em construir redes que modelassem como a memória pode ser armazenada. Vejamos os dois exemplos abaixo:



Suponhamos que, em ambos os casos, a célula 1 responda a um certo tipo de som e a célula S represente a memorização desse som.

Em (a), cada neurônio tem limiar 3 e o neurônio S fica ativado (armazenando a memória do som) somente após o som ter sido ouvido por três unidades de tempo consecutivas.

Em (b), o neurônio S tem limiar 1 e fica ativado (armazenando a memória do som) se o som tiver sido ouvido em qualquer tempo passado, mesmo que no tempo presente ele não seja ouvido. Neste caso, o mecanismo de memorização é *reverberatório*.

Note que, no caso (b), não existe um mecanismo de contagem ou estimação do tempo. Segundo McCulloch e Pitts, “a atividade regenerativa dos círculos constituintes faz referência indefinida ao tempo passado”.

Aprendizado Hebbiano

Dois dos fenômenos cognitivos mais estudados pelos psicólogos são o aprendizado e a memória. Alguns dos primeiros experimentos que permitiram uma abordagem do aprendizado em termos de noções sobre a memória foram os experimentos do fisiologista e psicólogo russo Ivan Pavlov (1849 – 1936) sobre condicionamento (hoje chamado de condicionamento clássico ou pavloviano).

Pavlov notou, em seus experimentos, que os cães quase sempre salivam quando vêem comida, mas raramente ou nunca salivam quando recebem um estímulo neutro, como o som de um sino por exemplo. Ele chamou a visão da comida de estímulo não-condicionado e a salivação do cão ao ver a comida de resposta não-condicionada, porque esta não é uma resposta aprendida.

Pavlov repetiu várias vezes um experimento em que um sino era tocado ao mesmo tempo em que se mostrava um prato de comida a um cão. Após várias repetições desse experimento, Pavlov notou que o toque do sino, sozinho, já era suficiente para fazer o cão salivar. Ele denominou o som do sino de estímulo condicionado e a salivação que ocorria após o toque do sino de resposta condicionada.

Em uma série de artigos e livros publicados entre 1939 e 1952, o psicólogo behaviorista estadunidense Clark Hull (1884 – 1952) propôs que os experimentos de Pavlov poderiam ser entendidos em termos de dois tipos de memória: memória de curta duração e memória de longa duração. Segundo Hull, após o término do experimento de condicionamento, a memória consciente do toque do sino teria desaparecido, pois o animal estaria concentrado em outras coisas. Porém, a memória da *associação sino-comida* ainda estaria presente, fazendo com que o cão salivasse prontamente após ouvir um novo toque do sino.

Hull fez a distinção entre traços de estímulo, sujeitos a um decaimento rápido, e forças associativas, capazes de persistir por um longo período.

Em termos do modelo de redes neurais de McCulloch e Pitts, poderíamos fazer uma analogia entre os traços de estímulo de Hull e as atividades (saídas) dos neurônios. A cada passo de tempo um neurônio fornece uma saída, mas no passo seguinte ele “a esquece” e fornece outra. Já as forças associativas poderiam ser entendidas caso o modelo de McCulloch e Pitts atribuísse “forças”, ou *pesos*, às conexões sinápticas entre os neurônios, as quais teriam valores que poderiam variar no tempo, mas com velocidades bem menores do que a definida pela mudança do passo de tempo do modelo. A idéia é que as sinapses teriam seus pesos alterados com a experiência do indivíduo, ou seja, com os novos aprendizados.

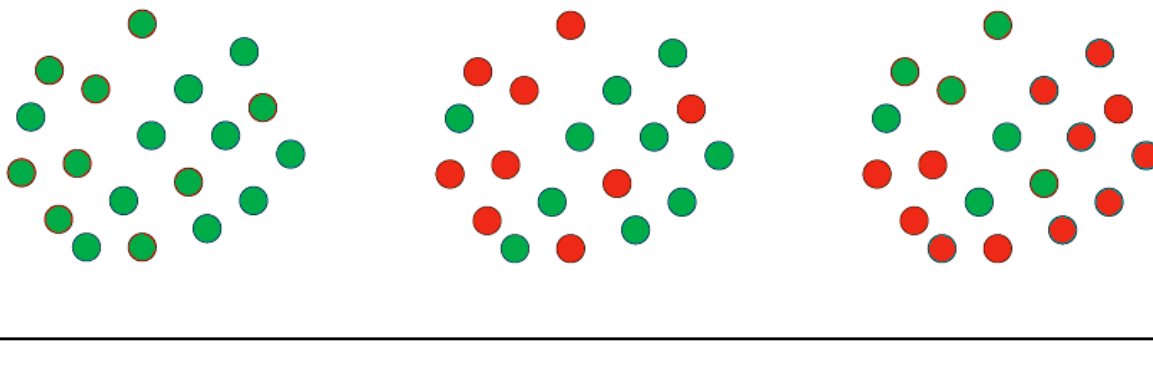
Hull chegou a propor várias equações matemáticas para representar a força associativa entre o estímulo condicionado e a resposta condicionada e como ela poderia variar (aumentar) a cada nova repetição do pareamento entre o estímulo não-condicionado e o estímulo condicionado. Os seus modelos, no entanto, não tiveram impacto duradouro (veja, porém, a aula em que falarmos sobre a regra de aprendizado de Rescorla-Wagner).

De impacto muito maior foi a proposta de mecanismo de memória feita pelo psicólogo canadense Donald Hebb (1904 – 1985) em seu livro de 1949 *The Organization of Behavior* (um dos livros mais citados de toda a história da psicologia e da neurociência).

Hebb discutiu o papel de circuitos reverberatórios no cérebro e argumentou que eles poderiam ser úteis como modelos de memória de curta duração, mas não como modelos de memória de longa duração. Segundo Hebb, sinais reverberatórios seriam instáveis e facilmente interrompidos por eventos externos (uma constatação que também foi feita por von Neumann em seus comentários sobre o cérebro e o computador). Para Hebb, uma memória de longa duração *estável* deveria depender de alguma *mudança estrutural* nas conexões (sinapses) entre os neurônios.

Segundo Hebb, durante a vida de um indivíduo, em função das suas experiências, certas conexões sinápticas entre grupos de neurônios tornam-se fortalecidas fazendo com que os circuitos formados por esses neurônios tornem-se por sua vez salientes, atuando como uma entidade única em meio ao vasto circuito neural cerebral. Hebb chamou tais circuitos neuronais de “assembléias neuronais” (*cell assemblies*) (veja a figura a seguir).

Assembléias neuronais (cell assemblies): grupos de células em vermelho



Para Hebb, as assembléias neuronais seriam as unidades *funcionais* básicas do cérebro. Segundo ele, uma assembléia neuronal poderia sustentar atividade cerebral autônoma mesmo na ausência de estímulos externos. A hipótese das assembléias neuronais seria, portanto, uma maneira de explicar como as memórias humanas se formam e são mantidas por longos períodos de tempo.

Nas palavras do próprio Hebb (op. cit. p. 62; os itálicos são do original):

“Existem memórias que são estabelecidas instantaneamente, e tão evanescentes quanto imediatas. ... Por outro lado, algumas memórias são ambas estabelecidas instantaneamente e permanentes. Para dar conta da permanência, alguma mudança estrutural parece necessária, mas um crescimento estrutural presumivelmente iria requerer um tempo apreciável. Se for possível encontrar alguma maneira de supor que um traço reverberatório possa cooperar com a mudança estrutural, e *manter a memória até que a mudança pelo crescimento ocorra*, poderíamos reconhecer o valor teórico do traço que é apenas uma atividade, sem precisar atribuir toda a memória a ele. O conceito de um traço reverberatório instável é, portanto, útil se for possível também supor que alguma mudança estrutural mais permanente o reforça. Não há razão para pensar que se deva escolher entre os dois conceitos; podem existir traços dos dois tipos, e memórias que dependam de ambos”.

Na sequência deste trecho, Hebb formula, como “Postulado Neurofisiológico”, a hipótese que deu origem ao que hoje se chama de sinapse hebbiana (op. cit., p. 62; a hipótese está em itálico, como no original).

“Vamos assumir então que a persistência ou repetição de uma atividade reverberatória (ou “traço”) tenda a induzir mudanças celulares duradouras que aumentem sua estabilidade. A hipótese pode ser formulada explicitamente como segue”:

“Quando um axônio de uma célula A está próximo o suficiente de uma célula B para excitá-la e, repetida e persistentemente, toma parte em fazê-la disparar, algum processo de crescimento ou mudança metabólica ocorre em uma ou ambas as células de maneira que a eficiência de A, como uma das células que fazem B disparar, é aumentada”.

Hebb não propôs uma formulação matemática para esta hipótese. Em seu livro, ele apenas descreve um possível mecanismo de reforço da eficiência sináptica entre dois neurônios em que os botões sinápticos cresceriam, aumentando a área de contato sináptico entre as células. A figura a seguir tenta ilustrar essa idéia.

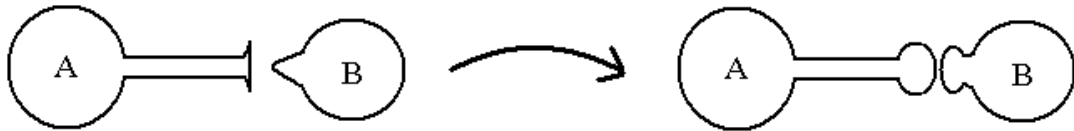
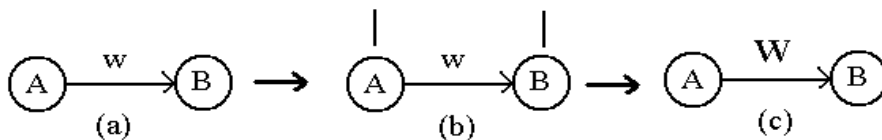


Diagrama do esquema proposto por Hebb para a mudança estrutural associada ao aumento da eficiência sináptica.

Um processo como o proposto por Hebb, de crescimento dos botões sinápticos, ainda não foi observado em animais adultos. Porém, outro mecanismo fisiológico foi observado no cérebro que parece ser uma evidência da existência de um processo de aumento da eficiência sináptica do tipo proposto por Hebb. Este mecanismo, conhecido como potencialização de longa duração (ou LTP, da sigla em inglês para *long-term potentiation*), foi observado inicialmente no hipocampo em 1973. Para uma revisão sobre LTP – e também sobre depressão de longa duração (LTD, do inglês para *long-term depression*) – veja as referências no fim desta aula.

Independentemente do mecanismo fisiológico causador do aumento da eficiência sináptica proposta por Hebb, e até mesmo da existência de tal mecanismo, a hipótese de Hebb pode ser escrita em termos matemáticos. A maneira mais simples de expressar a hipótese de Hebb matematicamente é supor que existe um *peso* w associado a uma sinapse entre dois neurônios e que esse peso aumenta quando os dois neurônios estão ativos *ao mesmo tempo* (veja a figura abaixo).



Modelo matemático para uma sinapse hebbiana:

- (a) o neurônio A faz uma sinapse de peso w com o neurônio B.
- (b) quando o neurônio A está ativo, o neurônio B está ativo ao mesmo tempo.
- (c) a ativação simultânea dos neurônios A e B provoca um aumento do peso da sinapse.

Uma expressão matemática que implementa este processo é:

$$\Delta w_{BA} = \eta x_B x_A,$$

onde w_{BA} é a peso da sinapse que A faz em B, x_A é a atividade de A, x_B é a atividade de B e η é o “parâmetro de aprendizagem”, uma constante positiva que indica de quanto w aumenta quando A e B estão ativos.

Segundo esta expressão, se $x_A = x_B = 1$, $\Delta w_{BA} = \eta.1.1 = \eta$. Já se $x_A = 1$ e $x_B = 0$, ou se $x_A = 0$ e $x_B = 1$, ou ainda, se $x_A = x_B = 0$, $\Delta w_{BA} = 0$.

Essa formulação matemática da hipótese de Hebb deixa clara a sua natureza: a eficiência sináptica, representada pelo peso w , aumenta em função da **correlação temporal** entre as atividades dos neurônios A e B.

A hipótese de Hebb pode ser expressa matematicamente de outras maneiras diferentes da mostrada acima. Considera-se, no entanto, que para que um modelo de sinapse possa ser chamado de hebbiano, ele deve possuir os seguintes mecanismos (Brown *et al.*, 1990):

1. Mecanismo de dependência temporal: a modificação em uma sinapse hebbiana deve depender do *tempo exato* de ocorrência dos sinais pré- e pós-sináptico.
2. Mecanismo local: a modificação em uma sinapse hebbiana deve ser uma função apenas da informação disponível *localmente* na junção sináptica.
3. Mecanismo interativo: a modificação em uma sinapse hebbiana depende dos sinais nos *dois lados* da sinapse (pré e pós).
4. Mecanismo correlacional: a modificação em uma sinapse hebbiana é causada pela *correlação temporal* entre os sinais pré- e pós-sináptico. Obviamente, não pode haver simultaneidade exata entre os sinais pré- e pós-sináptico. Devemos imaginar como correlacionados dois sinais que ocorram dentro de um pequeno intervalo de tempo, sendo que, para que a *causalidade* seja satisfeita, o sinal no neurônio pré-sináptico deve ocorrer antes do sinal no neurônio pós-sináptico.

Modernamente, a hipótese de Hebb foi estendida para incluir o caso em que o neurônio pré-sináptico A fica ativo depois do neurônio pós-sináptico B.

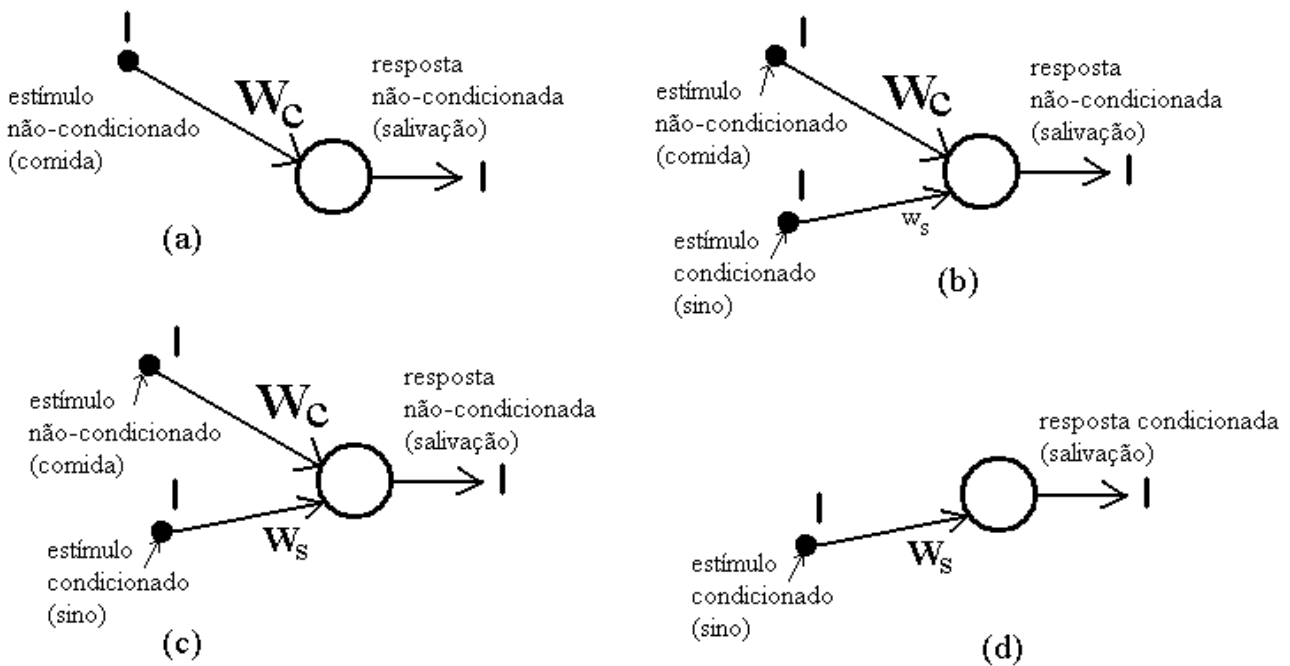
Neste caso de assincronia temporal, ou de falta de causalidade entre a atividade do neurônio pré-sináptico e a atividade do neurônio pós-sináptico, o peso da sinapse deve diminuir (como em uma LTD).

Outra maneira pela qual o peso de uma sinapse poderia diminuir, esta já imaginada por Hebb em 1949, seria no caso de uma sinapse não, ou pouco, usada.

Um fenômeno de alteração na eficiência sináptica (tanto de aumento como de diminuição) é chamado de **plasticidade sináptica**. A regra de Hebb, portanto, é uma regra de plasticidade sináptica. Como a hipótese de plasticidade sináptica implica, no fundo, em um modelo de aprendizado, a regra de Hebb é também chamada de *lei de aprendizado hebbiano*.

É curioso notar que o próprio Hebb ficou surpreso ao ver que o que ele chamou de “postulado neurofisiológico” em seu livro acabou sendo chamado pela comunidade científica de “regra de aprendizado hebbiana”. Para Hebb, esse postulado era um dos poucos aspectos da sua teoria que ele não considerava completamente original. Algo similar já havia sido proposto por outros psicólogos antes dele, incluindo Freud em seus primeiros anos como neurobiólogo (Milner, 1993, p. 127; disponível eletronicamente em formato pdf em <http://www.caam.rice.edu/~cox/worthing/milnersciam.pdf>).

Como aplicação simples da regra de Hebb, mostra-se a seguir que ela é suficiente para modelar o fenômeno de condicionamento clássico, observado por Pavlov (veja a próxima figura).



Na figura acima, a seqüência de (a) a (d) ilustra os passos do experimento de condicionamento clássico de Pavlov. O peso da conexão entre o estímulo não-condicionado (comida) e a resposta não-condicionada (salivação) é muito grande, provocando a resposta quando ocorre o estímulo. Já o peso da conexão entre o estímulo condicionado (sino) e a unidade de salivação é pequeno, de maneira que o toque do sino não provoca resposta. A repetição do pareamento entre os estímulos não-condicionado e o condicionado cria uma correlação entre as unidades do sino e da salivação, que eleva o peso da sinapse entre elas. Após certo número de repetições, o peso é tão grande que apenas o toque do sino é suficiente para provocar a salivação.

Na década de 1950 começaram a surgir os primeiros trabalhos científicos em que o computador era usado para a construção de *simulações* de sistemas físicos, químicos, biológicos, etc. Naquela época, alguns grupos se interessaram em simular redes de neurônios de McCulloch e Pitts conectados entre si por sinapses hebbianas.

A primeira dessas simulações foi feita por Farley e Clark em 1954, usando um dos computadores então existentes no MIT (*Massachusetts Institute of Technology*).

Outra simulação foi realizada por Rochester, Holland, Haibt e Duda em 1956, num dos laboratórios de pesquisa da IBM. O computador utilizado por eles era um IBM 701, com 2 Kbytes de memória.

O objetivo de Rochester *et al.* era estudar o modelo de aprendizado proposto por Hebb. Na primeira simulação feita, os neurônios implementados eram do tipo de McCulloch e Pitts, ou seja, binários e com limiar, e a regra de mudança do peso da sinapse feita por um neurônio 1 num neurônio 2 era

$$\omega_{21}(t+1) = \omega_{21}(t) + \eta x_1(t)x_2(t),$$

onde a variável t denota o número de iterações. Note que ω_{21} só aumenta quando, num dado passo de tempo t , tanto a atividade do neurônio 1 (x_1) como a atividade do neurônio 2 (x_2) forem iguais a 1. O incremento dado a ω_{21} é η .

A primeira descoberta de Rochester *et al.* foi que, à medida que o tempo passa, essa lei leva os pesos das sinapses a crescer sem limites.

Para evitar isso, eles propuseram uma *regra de normalização*: a soma de todos os pesos sinápticos da rede é sempre igual a uma constante C ,

$$\sum_{i=1}^N \omega_i = C,$$

onde N é o número de sinapses na rede. Desta forma, o crescimento de uma sinapse implicaria no enfraquecimento de outra. Se, ao final de um passo de iteração, o peso de uma das sinapses (por exemplo, a sinapse k) fosse aumentado por uma quantia $\Delta\omega_k = \alpha$, antes que a próxima iteração fosse iniciada todos os pesos seriam alterados pela subtração de uma quantia igual a α/N de cada um deles.

Desta forma, o peso ω_k seria aumentado, de fato, por uma quantidade igual a

$$\Delta\omega_k = \alpha - \frac{\alpha}{N} = \frac{\alpha(N-1)}{N}$$

e todos os demais pesos seriam reduzidos de uma quantia α/N ,

$$\Delta\omega_j = -\frac{\alpha}{N} \quad (j \neq k).$$

Variantes desta regra continuam sendo usadas até hoje.

Outras modificações que Rochester *et al.* fizeram para melhorar suas simulações foram:

- Uso de sinapses inibitórias, mas sem o poder absoluto que elas tinham no modelo de McCulloch e Pitts.
- Pesos das sinapses variando continuamente entre -1 e 1 .
- Neurônios não-binários, com saída variando continuamente entre 0 e um valor máximo.

O próximo capítulo desta história das redes neurais e do conexionismo, que trata de um tipo de rede neural muito influente chamado de *perceptron*, será deixado para outra aula.

Referências:

1. Abbott, L.F. and Nelson, S.B., Synaptic plasticity: taming the beast. *Nature Neuroscience*, 3(suppl.):1178-1183, 2000.
2. Anderson, J. A. and Rosenfeld, E. (Eds), *Neurocomputing: Foundations of Research*, MIT Press, Cambridge, MA, USA, 1989.
3. Brown, T.H., Kairiss, E.W. and Keenan, C.L., Hebbian synapses: biophysical mechanisms and algorithms. *Annual Review of Neuroscience*, 13:475-511, 1990.
4. Malenka, R.C. and Bear, M.F., LTP and LTD: an embarrassment of riches. *Neuron*, 44:5-21, 2004.
5. Milner, P. M., The Mind of Donald O. Hebb. *Scientific American*, 268:124-129. 1993.
6. Sejnowski, T.J., The book of Hebb. *Neuron*, 24:773-776, 1999.
7. Squire, L. R., Berg, D., Bloom, F. E., du Lac, S., Ghosh, A. and Spitzer, N. E., *Fundamental Neuroscience*, Third edition. Burlington, MA: Academic Press, 2008.