

## Cap. XIII

# INTENCIONALIDADE E REPRESENTAÇÃO MENTAL

## A intencionalidade pode ser naturalizada? Há células da vovó?

### 1. Intencionalidade em Brentano

O termo “intencionalidade” ganhou destaque na Filosofia da Mente com o filósofo e psicólogo alemão Franz Brentano,<sup>169</sup> que argumentou que a “referência a um objeto” é uma característica dos fenômenos psíquicos que está ausente nos fenômenos físicos: “podemos definir os fenômenos psíquicos dizendo que são aqueles fenômenos que contêm em si, intencionalmente, um objeto” (BRENTANO, 1874, p. 116). Como sinônimos de “intencionalidade” (que ele chama de “inexistência intencional”), Brentano sugeriu “referência a um conteúdo, direção a um objeto (pelo qual não se deve entender uma realidade), ou objetividade imanente” (p. 115). Modernamente, utiliza-se também o termo “sobredade”, tradução de “*aboutness*”.<sup>170</sup>

Que tal objetividade é “imanente” fica claro quando representamos um objeto fictício, como um cavalo alado. A ideia de um cavalo alado refere-se, segundo Brentano, a um objeto imanente (em nossa mente, ou num espaço de projeção de nossa mente); analogamente, a ideia de um objeto real refere-se, antes de tudo, ao objeto imanente correspondente. Posteriormente, Brentano mudaria sua opinião sobre a existência desses objetos imanentes.

Um ponto discutido por Brentano, com sua representação interna do filósofo escocês William Hamilton, é se todos os estados mentais são intencionais. Brentano defende que sim, “na ideia há algo ideado; no juízo há algo admitido ou recusado; no amor, amado; no ódio, odiado; no apetite, apetecido, etc” (p. 115). Mas Hamilton negou que sentimentos (*feelings*), incluindo o prazer e a dor, se refiram a algum objeto.<sup>171</sup>

De qualquer maneira, há um acordo de que estados envolvendo “atitudes proposicionais” (termo cunhado por Russell em 1918) são intencionais, expressos por proposições como “creio que...”, “desejo que...”, “imagino que...”, “duvido que...”, etc.

Independentemente de se a intencionalidade é uma propriedade de todos os estados mentais, ou só de parte deles, o desafio lançado ao fisicismo é de que “nenhum fenômeno físico revela algo parecido” (p. 126). Se é assim, como o fisicismo pode pretender reduzir estados mentais a estados físicos?

---

169 BRENTANO, F. (1874), *Psychologie vom empirischen Standpunkt*. Leipzig: Duncker & Humblot, Livro II, Cap. I. Em aula entregamos uma tradução dos § 5 e 9, pp. 115-8, 126-7. Em inglês: *Psychology from an empirical standpoint*, editado por Linda L. McAlister, trad. A.C. Rancurello, D.B. Terrell & L.L. McAlister, 2ª ed., 1995, Routledge, London. Na internet há também a tradução espanhola: *Psicología desde un punto de vista empírico*, trad. H. Scholten.

170 No termo “inexistência intencional de um objeto” (*intentionale Inexistenz eines Gegenstandes*), “o prefixo ‘in-’ não indica negação, mas localização, indicando existência na mente”, p. 423 de MCALISTER, Linda L., 1970, “Franz Brentano and intentional inexistence”, *Journal of the History of Philosophy* 8: 423-30. O termo “sobredade” é usada na tradução do livro didático de MASLIN, K.T. (2009), *Introdução à filosofia da mente*, 2ª ed. (de 2007, em inglês), trad. Fernando J. R. da Rocha, Artmed, Porto Alegre, p. 20, livro este que faz uma introdução muito boa à questão da intencionalidade na Filosofia da Mente.

171 HAMILTON, W.(1860), *Lectures on metaphysics and logic*, W. Blackwood, Edinburgh, 4 vols. Publicação póstuma organizada por H.L. Mansel & J. Veitch, redigida originalmente em 1836-8. Ver vol. II (não I, como indicado por Brentano), p. 432. (Perdoem a piadinha da representação interna.)

## 2. Representação mental

A intencionalidade é a propriedade que uma *representação* tem de se referir a alguma coisa, ou no mundo exterior (incluindo neste caso ocorrências internas também, como ocorre na introspecção) ou como um “objeto imanente”, segundo a concepção de Brentano. Mas o que é uma representação mental?

Façamos um exercício: pense em sua habitação, e diga quantas janelas há nela. Geralmente conseguimos contar e dizer o número correto, e isso só pode ser explicado supondo que temos em nossa mente uma representação mental de nossa habitação. A hipótese de que há representações em nossa mente explica muito bem o nosso comportamento e nossa adaptação ao ambiente. No exemplo dado, pode-se dizer que a representação surgiu com nossa experiência da habitação, ou seja, há geralmente um *processo causal* partindo do representado para a representação. Outra característica da representação é que ela pode ser falsa ou *errônea*.

Há muito debate filosófico sobre a natureza da representação mental, e não temos condições de fazer jus a ela no presente momento. Um ponto a ser mencionado são as chamadas “guerras da representação”, em que se discute a natureza da representação na Psicologia Cognitiva, e nas quais diferentes variantes do “representacionismo” são contestadas pelo movimento 4E (que mencionamos na seção XII.1) que defende que “o conceito de representação interna deve ser marginalizado ou até eliminado das ciências da mente e do comportamento”<sup>172</sup>, uma visão às vezes chamada de “dinamicista”.

Já dentro do representacionismo, um célebre debate foi se nossas representações são fundamentalmente proposicionais (Pylyshyn) ou se são imagéticas (Kosslyn).<sup>173</sup>

## 3. Representação na neurociência

No contexto da neurociência, desde o início do século XX conhecem-se áreas, como o córtex somatossensorial, que possuem partes que são ativadas especificamente por um estímulo em algum ponto do corpo (ver seção I.7). Isso ocorre também no caso do córtex visual primário, exibido na Fig. XI.1, onde ocorre uma cópia deformada de um objeto externo. Podemos falar em um “isomorfismo”, ou simplesmente um “mapeamento”, de pontos do mundo externo em pontos da representação.<sup>174</sup>

Há uma tendência de dizer que os córtices somatossensorial e visual “representam” o corpo, mas se aceitarmos que a consciência emerge em partes mais profundas do encéfalo (ver seção IX.5), então estas áreas primárias do córtex não seriam propriamente uma “representação *mental*”, mas talvez uma “representação intermediária”.

Por outro lado, o córtex *motor* também tem uma relação de isomorfismo com o corpo, mas neste caso o processo causal parte do estímulo do córtex e vai em direção ao movimento

---

172 WILLIAMS, D. (2018), “Predictive processing and the representation wars”, *Minds & Machines* 28: 141-72, citação das pp. 142-3. Autores mencionados na tradição dinamicista do 4E são Varela, Thompson & Rosch (1993), Chemero (2009), Lawrence Shapiro (2009), Hutto & Myin (2013) e Michael Anderson (2014).

173 KOSSLYN, S.M. (1981), “The medium and the message in mental imagery: a theory”, *Psychological Review* 88: 46-66. PYLYSHYN, Z.W. (1981), “The imagery debate: analogue media versus tacit knowledge”, *Psychological Review* 87: 16-45. TYE, M. (1991), *The imagery debate*, MIT Press, Cambridge.

174 É costume falar-se em “isomorfismo estrutural”, como em POLDRACK, R.A. (2020), “The physics of representation”, *Synthese*, pp. 1-21, mas tal uso não parece muito rigoroso.

do corpo. Para diferenciar estas situações, pode-se dizer que o córtex somatossensorial “representa fatos” e que o córtex motor “representa metas”.<sup>175</sup>

Um exemplo de representação em área cerebral mais profunda são as “células de lugar” (*place cells*) no hipocampo, cada uma das quais responde mais fortemente em um determinado local do mundo externo.

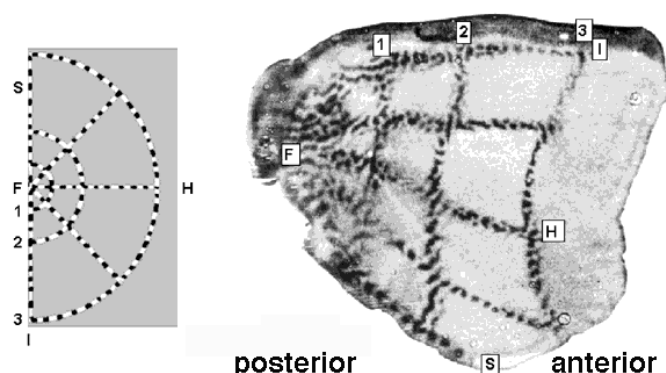


Figura XIII.1. Mapa retinotópico na área visual primária VI em macaco rhesus, obtido a partir do padrão à esquerda. Áreas estimuladas consomem mais glicose, associada a carbono-14 radioativo, que aparece em um filme fotográfico após a remoção do tecido cerebral (TOOTELL et al., 1998, p. 902).<sup>176</sup>

A existência de uma correlação e de uma relação causal entre dois domínios não é suficiente para haver “representação”. Por exemplo, as células beta do pâncreas respondem a variações do nível de açúcar no sangue, mas não diríamos que eles “representam” o nível de açúcar (POLDRACK, 2020, p. 13). Se um médico monitorar as células beta de um paciente, ele poderia considerar que elas estão representando o nível de açúcar, mas isso se dá no contexto de sua própria mente interagindo com a informação do sistema; sem esta mente acoplada ao sistema, não diríamos que há “representação”. Isso sugere que a noção de representação mental deve desempenhar uma função, função esta que não é desempenhada por um mero “receptor” como a célula beta, e que está atrelada à ação do organismo no ambiente externo. Ou então, que uma representação deve fazer parte de um “sistema representacional”, como o sistema nervoso (fica em aberto a questão de se o sistema imunológico envolve representações). Sobre a questão de se um robô “representa” seu ambiente, tal questão não deve ser separada do fato de que ele foi construído por seres humanos imbuídos de representação mental, e que a representação que a máquina teria está acoplada de alguma maneira com a representação mental dos seus construtores.

Esboçemos agora algumas condições necessárias para se estabelecer uma representação.

(i) Há de se definir um *veículo da representação*, ou seja, uma realização física que representa algum outro domínio da realidade. Tal veículo funciona segundo leis causais internas ao sistema. É plausível supor que mecanismos causais de retroalimentação (*feedback*, ou processamento recorrente) sejam essenciais para o estabelecimento de representações.<sup>177</sup>

175 MILLIKAN, R.G. (2000), “Naturalizing intentionality”, in Elevelitch, B. (org.), *Philosophy of mind*, vol. 9 de *The proceedings of the Twentieth World Congress of Philosophy*, Philosophy Documentation Center, Bowling Green (OH), pp. 83-90; republicado em Carmo, Juliano S. do (org.) (2015), *A companion to naturalism*, Dissertação Filosofia, NEPFIL-UFPEL, Pelotas, pp. 211-16.

176 A imagem da Fig. XIII.1 foi obtida por TOOTELL, R.B.H.; SILVERMAN, M.S.; SWITKES, E. & DE VALOIS, R.L. (1982), “Deoxyglucose analysis of retinotopic organization in primate striate cortex”, *Science* 218: 902-4.

177 Na literatura recente, LAMME, V.A.F. (2006), “Towards a true neural stance on consciousness”, *Trends in Cognitive Sciences* 10: 494-501, é um que considera que a retroalimentação é necessária para a experiência consciente (p. 497).

(ii) Deve haver uma *correlação* ou isomorfismo entre a representação e o representado. É a partir dessa correlação que se define, em termos filosóficos, o “conteúdo” da representação, ou seu aspecto semântico.

(iii) Há uma *relação causal* partindo do mundo externo representado e indo em direção à representação interna, mesmo que tal relação seja bastante indireta, como em representações mais abstratas.

(iv) Representação não é mera “recepção”, ou seja, não basta haver uma correlação causal entre dois sistemas. Se dissermos que um neurônio “representa”, *ele deve fazer parte de um sistema representacional complexo*, como o sistema nervoso. Tal sistema deve ter uma função específica de ação no ambiente, que no caso de seres vivos surgiu através de um mecanismo de seleção natural.

(v) Uma representação pode ser classificada de *errônea*, ou seja, pode romper a correlação esperada entre representação e mundo representado. Um exemplo simples são as representações mentais em um sonho.

(vi) A distinção entre as duas direções (sentidos) de causalidade deve ser analisada no sistema sob estudo. Além da “representação de fatos” enfatizada até aqui, há também a “representação de metas”, em que uma representação imaginada leva a uma ação sobre o mundo, tendo em vista ajustá-lo à representação. Em seres humanos, temos aqui todo o vasto domínio da *imaginação*, que envolve manipulação de representações.

Por fim, poder-se-ia exigir que uma “representação” pressupusesse um sistema consciente, ou uma mente (ou seja, exigir que toda representação fosse uma representação mental). Mas estamos interessados em uma definição mais ampla de representação, que em princípio poderia existir em um sistema complexo que tenha evoluído por seleção natural, em um planeta em que os seres não são conscientes.<sup>178</sup>

Para explorar os itens (iv), (v) e (vi), visitaremos na seção XI.5 a área filosófica conhecida como “teleossemântica”, que busca naturalizar a intencionalidade.

#### 4. O que o olho da rã diz para seu encéfalo?

Um estudo pioneiro na neurofisiologia foi feito por Jerome Lettvin, em colaboração com Humberto Maturana (que posteriormente formulou a teoria da autopoiesis) e com a dupla Warren McCulloch & Walter Pitts, que foram pioneiros na interpretação de neurônios como porta lógicas em 1943.<sup>179</sup> Eles estudaram a visão em uma rã, e identificaram quatro padrões que o olho da rã informa para seu cérebro por meio do nervo óptico (nenhum dos quais é uma informação sobre a iluminação em células individuais da retina): “1) bordas nítidas locais e contraste; 2) a curvatura de uma borda de um objeto escuro; 3) o movimento das bordas; e 4) a diminuição local de iluminação provocada por movimento ou rápido escurecimento geral” (p. 1951). Descobriram que cada grupo de fibras implementando essas operações gera um mapa em uma folha individual de terminações no colículo superior da rã.

As operações têm assim muito mais o sabor de percepção do que de sensação, se é que tal distinção tem qualquer significado atualmente. Ou seja, a linguagem em que elas são melhor descritas é a linguagem de abstrações complexas a partir da imagem visual. Estamos tentados, por exemplo, a chamar os detectores de convexidade de “percebedores

178 Como exemplo, ver o cenário montado em NERI, H. & PESSOA JR., O. (2018), “Science without consciousness”, in Chibeni, S.S. et al. (orgs.), *Filosofia e historia de la ciencia en el Cono Sur: selección de trabajos del X Encuentro*, AFHIC, Córdoba (AR), pp. 149-55.

179 LETTVIN, J.; H. MATURANA, H.; MCCULLOCH, W. & PITTS, W. (1959), “What the frog’s eye tells the frog’s brain”, *Proceedings of the IRE* 47: 1940-1951. O termo “bug detector” fora usado por Horace Barlow, em 1953.

de bichinhos” [“*bug perceivers*”]. Tal fibra (operação 2) responde melhor quando um objeto escuro, menor que o campo receptivo [região da retina que faz disparar a célula ganglionar em questão], entra naquele campo, para, e depois se move intermitentemente. (LETTVIN et al., 1959, p. 1951)

A sugestão é que essas fibras ópticas da operação 2 permitem a “representação” de objetos convexos escuros, que no contexto natural são geralmente “bichinhos escuros” (insetos ou minhocas).

## 5. Naturalização da intencionalidade: teleossemântica

A expressão “naturalização da intencionalidade” refere-se à tentativa de mostrar como a intencionalidade pode ser explicada a partir das ciências naturais. No contexto contemporâneo, tal projeto foi iniciado por Jerry Fodor (1975), que defendeu que uma representação faz parte do mundo material e de suas cadeias causais. Porém, ficou em aberto a questão de qual propriedade física relacional complexa seria a intencionalidade, que permitiria que uma representação pudesse ser “errônea”. Fred Dretske (1981) também atacou o problema de uma perspectiva informacional, mas sua teoria não é considerada plenamente satisfatória. Dois filósofos mais recentes são apontados como dando passos mais decisivos para o projeto de naturalização da intencionalidade conhecido como “teleossemântica”: a estadunidense Ruth Millikan (1984) e o inglês David Papineau (1984).<sup>180</sup>

Iniciaremos com o relato de SHEA (2006, 2013) sobre a concepção de Millikan. A *teleossemântica* tem duas partes principais. A primeira é salientar como a representação é *usada* pelo organismo em questão. A segunda é considerar a origem biológica da representação, ou seja, a origem da função desempenhada (SHEA, 2013, p. 501).

A função da percepção e da cognição é representar o ambiente. A intencionalidade é uma propriedade apenas de processos com um tipo especial de função: a função biológica de representação (SHEA, 2006, p. 4). Dado que uma das funções de um sistema é produzir representações, tais representações devem funcionar como representação para outra parte do próprio sistema.

Assim, uma parte do sistema *produz* a representação e outra parte a *consome* (MILLIKAN, 1989, pp. 283-5). Esta distinção é própria de Millikan, não sendo salientada por Papineau: a representação não é uma função de um sistema indiferenciado, mas envolve um sistema composto de produtor e consumidor, sendo que as representações (ou os atos associados) são intermediários produzidos por um e usados pelo outro, de forma que o sistema como um todo atue da maneira que foi selecionada naturalmente. Vejamos na Tabela XIII.1 alguns exemplos de produtor, ato da representação e consumidor (extraída de SHEA, 2006, p. 8), onde adicionamos sugestões para o “conteúdo” da representação.

---

180 Seguimos aqui SHEA, N. (2006), “Millikan’s contribution to materialist philosophy of mind”, *Matière Première 1*: 127-156 (preprint disponível online, pgs. 1-22). SHEA, N. (2013), “Naturalising representational content”, *Philosophical Compass 8*: 496-509. MILLIKAN, R.G. (1984), *Language, thought and other biological categories*, MIT Press, Cambridge. PAPINEAU, D. (1984), “Representation and explanation”, *Philosophy of Science 51*: 550-72. MILLIKAN, R.G. (1989), “Biosemantics”, *Journal of Philosophy 86*: 281-297. MILLIKAN (2000), op. cit. (nota 156). Este assunto é pesquisado pelo brasileiro Sérgio Farias de Souza Filho.

<i>Produtor</i>	<i>Ato da representação</i>	<i>Consumidor</i>	<i>Conteúdo</i>
Castor que bate seu rabo ao detectar um perigo	O bater do rabo na água	Outros castores: mergulhar para dentro d'água	Animal perigoso presente.
Abelha que dança ao retornar de campo de flores com néctar	A dança da abelha	Outras abelhas: voar na direção e distância indicadas	Campo de flores localizado.
Olho da rã e seu sistema óptico	Disparo da célula ganglionar da retina	Mecanismo de lançamento de língua da mesma rã	Bichinho escuro numa certa posição.

*Tabela XIII.1: Exemplos de representação no reino animal: produtor, consumidor, o ato da representação e uma interpretação nossa sobre seu conteúdo (adaptado de SHEA, 2006, p. 8).*

Vimos na seção anterior os estudos do sistema visual de rãs na década de 1950, em que se descobriu que as células ganglionares da retina respondiam seletivamente a pequenos pontos pretos movendo-se através do campo visual. Logo descobriu-se que essas células fazem disparar o reflexo de captura de moscas da rã. O reflexo é estimulado mesmo quando as células são ativadas artificialmente por meio de uma corrente elétrica. Assim, concluíram que essas células ganglionares *representam* a presença de um bichinho escuro, sendo que “bichinho escuro” devidamente localizado no campo visual seria o conteúdo da representação. Esse é um bom exemplo de como o conteúdo da representação está relacionado com o mecanismo de consumo (SHEA, 2006, p. 8).

Nos exemplos do castor e da abelha, o produtor e consumidor operam em organismos diferentes, mas no caso da rã e em muitos exemplos com humanos, os dois sistemas encontram-se no mesmo organismo, e as representações intermediárias são internas. Esta distinção permitiu que Millikan colocasse o foco no papel importante do mecanismo de consumo para determinar o conteúdo das representações intermediárias, a ser chamado de “abordagem de olhar para frente” (*forward-looking approach*) (SHEA, 2006, p. 7).

O texto “Naturalizando a intencionalidade” de Ruth Millikan (2000) levanta mais alguns pontos interessantes, mesmo sendo um texto difícil. Ela começa mencionando que, para Brentano, a intencionalidade envolve a capacidade de manter uma relação real com algo não existente, como um fato falso (item v da seção XIII.3). A esta relação problemática ela chama de “relação de Brentano” [*Brentano's relation*].

O primeiro ponto de Millikan é que a relação de Brentano ocorre não só para representações mentais, mas para qualquer propósito *não realizado*. Um exemplo de um propósito natural não-mental é o propósito do estômago de digerir comida, e se por algum motivo não ocorrer a digestão, o estômago deixa de cumprir sua função. Um estômago então mantém a relação de Brentano, apesar de não ter intencionalidade. Assim, intencionalidade implica a relação de Brentano, mas não o contrário.

Millikan (1984) definiu um “propósito natural” como uma *função própria*, no sentido de ser aquilo que levou à seleção natural dos antepassados do organismo. No caso mais simples, “as funções próprias de uma coisa são os efeitos que, no passado, deram conta da seleção de seus ancestrais por reprodução” (MILLIKAN, 2000, p. 85). Pode ocorrer que uma parte de um ser vivo não mais desempenhe a função original para a qual evoluiu: isso constitui uma versão naturalizada da relação de Brentano. Estados internos, como os *estados perceptivos e cognitivos* de organismos, têm funções próprias que variam conforme as entradas ambientais para tais sistemas geneticamente programados. Mas uma função natural só é associada à intencionalidade *quando o propósito é representado explicitamente*.

Segundo MILLIKAN (2000, p. 86), há duas variedades fundamentais de intencionalidade, duas direções de aptidão (*directions of fit*) para entidades intencionais: a direção que representa metas (*goal-representing direction*; representação imperativa, segundo SHEA, 2013, p. 510) e a direção que representa fatos (*fact-representing direction*; representação indicativa). A primeira, ligada à ação, seria mais fundamental. Quando um propósito natural é representado, estabelece-se uma correlação entre a representação e um estado de coisas externo, sendo que a função própria é realizar esse estado de coisas a partir de mecanismos cooperativos. Explica-se assim a intencionalidade de propósitos explicitamente representados. A partir disso, pode-se abstrair a intencionalidade de palavras e conceitos, que constituiriam a representação de fatos.

Eis então um esboço do projeto de naturalização da intencionalidade, proposto pela teleosemântica. Ele é um tanto complicado, então é preciso refletir com calma sobre suas teses, para avaliar sua plausibilidade.

## 6. Células da vovó

Jerome Lettvin, autor principal do artigo sobre representação em rãs (seção XI.4), gerou uma história engraçada, quando professor no MIT, em torno de 1969, ao discutir se neurônios podem representar objetos individuais. Inventou a história do neurocirurgião russo Akakhi Akakhievitch, que teria encontrado um grupo de 18 mil neurônios que só respondiam a “mãe”, “quer animada ou de pelúcia, vista de frente ou de trás, de cabeça para baixo ou em diagonal, ou mostrada em caricatura, fotografia ou abstração”. Nisso, o personagem principal do romance *O complexo de Portnoy*, de Philip Roth, obcecado por sua mãe, teria procurado o neurocirurgião para remover todas as células que representavam sua mãe. O cirurgião teve sucesso com isso, e o jovem perdeu completamente a memória de sua mãe, apesar de lembrar de um vestido vermelho vagando pela casa de chinelos, e de alguém gritando com ele quando namorava moças não judias. Vislumbrando o Prêmio Nobel, Akakhievitch teria se voltado então para pesquisar a “célula da vovó” (*grandmother cell*).<sup>181</sup>

Essa hipótese, de que a representação de um objeto particular do mundo estaria associada a um pequeno grupo de neurônios, e que geralmente não é compartilhada pela maioria dos neurocientistas, ficou conhecida como “teoria da célula da vovó”. No entanto, ela teve seus defensores e foi ressuscitada no século XXI, como veremos na seção seguinte. A hipótese tradicionalmente mais favorecida é a “codificação por população de células” (*ensemble coding*), em que uma vovó é representada por um distintivo padrão de atividade em uma população de neurônios. Essa representação distribuída é exemplificada no processamento paralelo em redes neurais artificiais.

Como explica o neurocientista e historiador Charles GROSS (2002), a teoria do neurônio da vovó foi defendida explicitamente pelo neurofisiologista polonês Jerzy Konorski, em 1967, no seu livro *Integrative activity of the brain*, recebendo o nome de “células gnósticas”. A ideia de que existem representações explícitas expressas em neurônios específicos já tinha sido antecipada nos experimentos de David Hubel & Torsten Wiesel, ganhadores do Nobel em 1981. Em 1959, eles descobriram células no córtex visual de gatos que respondiam a bordas locais com uma orientação específica. Horace Barlow, bisneto de Charles Darwin, se firmou como expoente desta escola de pensamento ao postular cinco dogmas da doutrina neuronal (BARLOW, 1972). Muitos experimentos desde então aumentaram o catálogo de células com respostas específicas e invariantes, como a resposta a imagens de mãos e rostos de macacos (*face*

---

181 GROSS, C.G. (2002), “Genealogy of the ‘grandmother cell’”, *Neuroscientist* 8: 512-8. As citações sobre Akakhievitch são transcritas por Gross de carta de Lettvin a Barlow em 1969. BARLOW, H.B. (1972), “Single units and sensation: a neuron doctrine for perceptual psychology?”, *Perception* 1: 371-94. JAMES (1890), op. cit. (nota 102).

*neurons*), a ativação por movimentos específicos próprios ou de outro macaco, importantes para o estudo da interação social (neurônios espelho), ou neurônios do hipocampo de um rato, ativados apenas quando este está numa posição específica no ambiente (*place cells*).

GROSS (2002, p. 517) salienta ainda duas características de células da vovó. A primeira é chamada de “codificação de linha rotulada” (*labeled line coding*), quando um neurônio codifica uma propriedade particular de estímulo. Isso é exemplificado pelos achados de Johannes Müller (1835) de que um nervo reage sempre da mesma maneira a um estímulo qualquer em um receptor sensorial específico. Codificação de linha rotulada para um único neurônio foi observado pela primeira vez por Edgar Adrian & Bryan Matthews (1927), ao mostrar que os potenciais de ação em uma fibra nervosa do congro (enguia-do-mar) correspondiam à estimulação óptica de uma parte específica da retina do animal.

A segunda característica é *convergência* de vários sinais para um único neurônio ou grupo de neurônios. Esta noção apareceu no curioso conceito de “célula pontifical” de William JAMES (1890, p. 179), que seria uma célula na qual nossa consciência estaria localizada (um “sensório comum” reduzido a uma única célula). James rejeitou esta ideia, em sua crítica a uma versão do monadismo leibniziano, segundo a qual cada célula teria sua própria consciência.

## 7. Representação de conceitos em neurônios

Em 2005, o neurocientista argentino Rodrigo Quiroga e colegas, no laboratório de Christof Koch na Caltech, realizaram experimentos que reavivaram a hipótese da célula da vovó.<sup>182</sup> Pacientes com epilepsia podem ser submetidos a implantes de eletrodos profundos, e esses foram utilizados para medir a atividade de neurônios de uma região que inclui o hipocampo, e está relacionada a várias funções, incluindo a formação de novas memórias. Dezenas de imagens foram mostradas aos pacientes. Os neurônios desta área têm resposta esparsa, estando a maioria do tempo em silêncio, mas parte desses se excitaram quando alguma foto específica, ou fotos de um objeto específico, foram exibidas. Por exemplo, um grupo de neurônios respondeu apenas a uma foto da atriz Jennifer Aniston. A surpresa maior veio na segunda sessão, quando fotos adicionais relacionadas a Jennifer foram mostradas, incluindo fotos da atriz com diversas roupas, em diversas posições. Uma célula se ativou para todas essas fotos, e inclusive quando apenas o seu nome impresso foi mostrado, e não respondeu à foto de nenhuma outra atriz. Curiosamente, na foto em que Jennifer estava com Brad Pitt, esse neurônio não se ativou. Esse neurônio foi batizado de Jennifer Aniston. Foram encontradas células para Bill Clinton, torre de Pisa, animais pequenos fofos e mesmo para o teorema de Pitágoras (Fig. XI.2).

Experimentos seguintes mostraram ainda que essas células se formam com relativa rapidez, em poucos dias. Células ativadas exclusivamente na presença dos próprios pesquisadores e das enfermeiras, pessoas que eram anteriormente desconhecidas aos pacientes, foram logo encontradas.

O que podemos concluir? Que temos neurônios que respondem de forma altamente específica a certos objetos do mundo, de forma invariante a como esses objetos são apresentados e ao contexto em que se encontram. É como se houvesse um conjunto específico de neurônios relacionado ao conceito de um objeto particular ou classe de objetos. Com este argumento, os autores nomearam estes neurônios de “células de conceito” (*concept cells*);

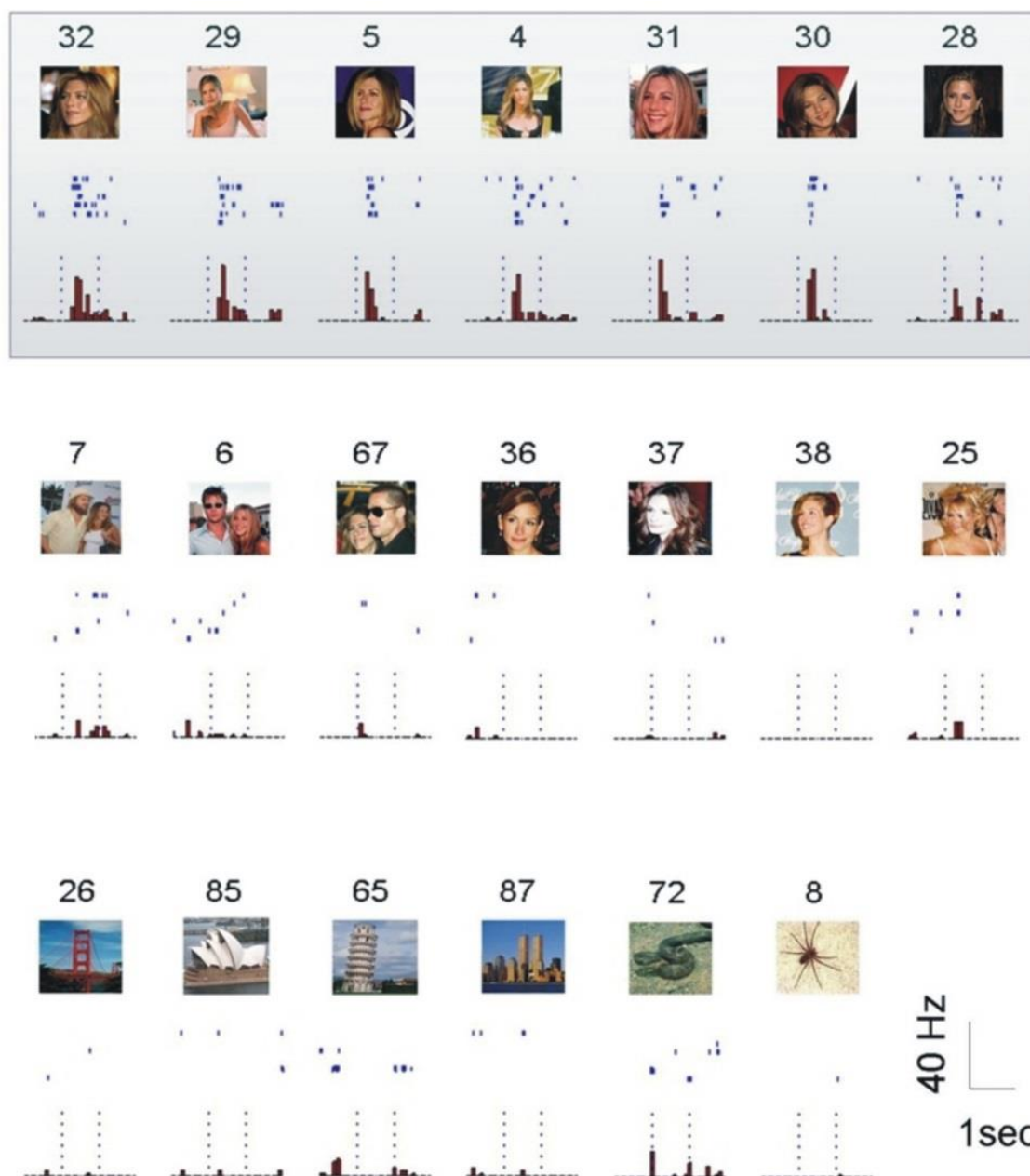
---

182 QUIROGA, R.Q.; REDDY, L.; KREIMAN, G.; KOCH, C.; FRIED, I. (2005), “Invariant visual representation by single neurons in the human brain”, *Nature* 435: 1102-7. O presente capítulo é modificado de NAVES DE BRITO, Carlos S. & PESSOA Jr., O. (2014), “Neurociência: em busca da compreensão do cérebro e da mente”, in Freire Jr., O.; Greca, I.M. & El-Hani, C.N. (orgs.), *Ciências na transição dos séculos: conceitos, práticas e historicidade*, EDUFBA, Salvador, pp. 265-85.



informalmente, porém, ficaram conhecidos como “*Jennifer Aniston cells*”. O experimento não mostra que existe uma única célula ativada para cada objeto percebido; das muitas que o são, apenas uma foi registrada pelo experimento. Por outro lado, está claro que a ativação de uma célula de conceito requer anteriormente várias etapas de processamento cortical.<sup>183</sup>

Por fim, é bem plausível que um conjunto de neurônios conceitos esteja conectado a muitos outros, de forma homóloga a como os conceitos em nossa mente se conectam, o que abre o diálogo do assunto com filósofos e linguistas.



183 Modelos computacionais, como os do grupo de Tomaso Poggio, no MIT, implementam esse processo, nos quais, a cada uma das camadas, aumenta a especificidade dos neurônios ativados em relação a objetos percebidos, mas também aumenta sua invariância a detalhes, como a posição dos objetos. Ver: SERRE, T.; WOLF, L.; BILESCHI, S.; RIESENHUBER, M.; POGGIO, T. (2007), “Robust object recognition with cortex-like mechanisms”, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29: 411-26.

*Figura XI.2. Respostas para algumas das 87 imagens mostradas, relacionadas a uma das células registradas. Cada resposta aparece com seu número, a imagem, um gráfico de rasteio e histograma com marcações temporais. Os gráficos de rasteio, que aparecem logo abaixo das imagens, contêm pequenos retângulos que indicam a detecção de um potencial de ação, ao longo de alguns segundos, para seis corridas experimentais (para cada imagem). Abaixo de cada gráfico há um histograma, que indica a soma dos potenciais de ação que aparecem no gráfico acima, e se encontra a marcação do tempo. As imagens aparecem durante um segundo, e os instantes em que aparecem e desaparecem estão indicados pelas linhas verticais pontilhadas. Notar como há ativações significativas apenas para fotos de Jennifer Aniston, na fileira de cima, mas não quando ela está com seu marido Brad Pitt (imagens 7, 6, 67) (adaptado de QUIROGA et al., 2005).*