

Universidade de Brasília

Instituto de Ciências Humanas

Departamento de Filosofia

Programa de Pós-Graduação em Filosofia

Aluna: Ana Luísa Lamounier Costa

Orientador: Samuel Simon Rodrigues

Dissertação de Mestrado

Explicações em Neurociência:

o papel da causalidade em explicações mecanísticas e multiniveladas

Brasília, DF

2013

ANA LUÍSA LAMOUNIER COSTA

**EXPLICAÇÕES EM NEUROCIÊNCIA:
O PAPEL DA CAUSALIDADE EM EXPLICAÇÕES
MECANÍSTICAS E MULTINIVELADAS**

Dissertação de Mestrado com vista à obtenção de título de Mestre
junto ao departamento de Filosofia da Universidade de Brasília.

Orientador: Prof. Dr. Samuel José Simon Rodrigues

Brasília, 2013

ANA LUÍSA LAMOUNIER COSTA

**EXPLICAÇÕES EM NEUROCIÊNCIA:
O PAPEL DA CAUSALIDADE EM EXPLICAÇÕES
MECANÍSTICAS E MULTINIVELADAS**

Dissertação de Mestrado apresentada ao Departamento de Filosofia da Universidade de Brasília para a obtenção de título de Mestre.

Orientador: Prof. Dr. Samuel José Simon Rodrigues

Data de aprovação: 27 de maio de 2013

Banca examinadora:

Prof. Dr. Paulo César Coelho Abrantes

Prof. Dr. Valdir Filgueiras Pessoa

Prof. Dr. Samuel José Simon Rodrigues

AGRADECIMENTOS

Ao meu pai, por me mostrar a paixão pelo conhecimento;

À minha mãe, pelo exemplo de dedicação e competência inigualáveis;

Ao tio Simão, por me ensinar a duvidar;

À vovó Rosa, pelo suporte em todos os sentidos, por estar sempre lá quando precisei;

Ao Cassiano, pelo amor e amizade, a compreensão, os ouvidos atentos e a enorme paciência;

Ao Pedro e Henrique, por serem os melhores amigos-irmãos que já existiram;

Ao Prof. Samuel, pelas inúmeras discussões prolíferas, por ouvir e respeitar minhas opiniões e pela amizade construída nesses anos de orientação;

Ao professor Valdir Filgueiras Pessoa, por ter incitado meu interesse pela neurociência ainda durante a graduação e pelas excelentes críticas construtivas feitas na qualificação;

Ao professor Paulo Abrantes, pelas magistrais aulas do mestrado e pelas imprescindíveis críticas feitas na qualificação;

À Brigitte, o Chico e o Pisquilha, pela companhia incansável nas madrugadas de estudo (mesmo dormindo o tempo todo).

Resumo

A neurociência como uma ciência autônoma consolidou-se a partir de pouco antes da metade do século XX. Ela inclui conhecimentos provenientes de muitas áreas diferentes, incluindo a bioquímica, a biologia molecular, neurofisiologia, a ciência da computação e a psicologia comportamental. Os seus objetivos, problemas e a forma de elaborar explicações são distintos das outras áreas da biologia e das ciências biomédicas. Um modelo de explicação para a neurociência precisa estar preparado para lidar com essas questões. A mais expressiva abordagem explicativa para a neurociência, atualmente, é o modelo mecanístico-causal elaborado por Carl Craver em 2007. Segundo essa abordagem, explicações em neurociência descrevem mecanismos, ou seja, como partes se organizam e interagem para produzir um fenômeno. Essas partes pertencem a vários níveis de composição e relacionam-se causalmente umas com as outras. Além disso, a neurociência é unificada quando cientistas das diversas áreas que a compõem trabalham juntos para elaborar mecanismos. Esse modelo foi extensamente discutido nos últimos anos, tendo sido levantadas diversas críticas contra eles. Algumas delas são pertinentes e sugerem que ele sofra alguns ajustes. Ainda assim, ele sobrevive como o mais robusto modelo para explicações em neurociência existente. Esse trabalho apresenta-se como um estudo desse modelo, bem como das críticas feitas a ele.

Abstract

Neuroscience as an independent science was consolidated hardly before the first half of the twentieth century. It encloses knowledge from many different areas, including biochemistry, molecular biology, neurophysiology, computer science and behavioral psychology. Its purposes, its problems and how its explanations are formulated are distinct from other areas of biology and biomedical sciences. An explanatory model for neuroscience has to be prepared to deal with these questions. The most expressive account of explanations in neuroscience is currently the causal-mechanistic model formulated by Carl Craver in 2007. According to him, explanations in neuroscience describe mechanisms, in other words, it points out how parts organize themselves and interact to engender the phenomenon. These parts belong to various levels of composition and are causally related to one another. Furthermore, neuroscience is unified when scientists from different areas that compose it work together to develop mechanisms. This model was extensively discussed in the last years and several criticisms were raised towards it. Some of them are relevant and suggest some adjustments. Still, it remains as the soundest model for explanations in neuroscience nowadays. This work is presented as a review of this model, as well as the critiques worked out against it.

Sumário

Introdução	1
Capítulo 1 – explicação em ciência geral	4
1.1 O modelo nomológico-dedutivo	4
1.1.1 Visão Geral	4
1.1.2 O modelo	5
1.1.3 O problema dos enunciados-leis	6
1.1.4 Contraexemplos famosos ao modelo nomológico-dedutivo	7
1.1.5 O problema da simetria	8
1.1.6 O problema da relevância explicativa	9
1.2 Explicação estatística	10
1.2.1 O modelo estatístico-dedutivo	10
1.2.2 O modelo estatístico-indutivo	11
1.2.3 Objeções ao modelo estatístico-indutivo	12
1.2.4 O modelo da relevância estatística (RE)	13
1.3 Explicação causal	14
1.3.1 A concepção ôntica da explicação	15
1.3.2 O modelo mecânico-causal e a relevância explicativa	16
1.3.3 O modelo mecânico-causal e os sistemas complexos	17
1.3.4 O modelo das quantidades conservadas (QC)	17
1.4 Explicação como unificação	18
1.4.1 O problema das assimetrias	19
1.4.2 Unificação não explicativa	20
1.5 Explicação em medicina e ciências biomédicas	20
Capítulo 2 – explicação em neurociência	23
2.1 Visão geral	23

2.1.1	Explicações em neurociência descrevem mecanismos _____	24
2.1.2	Explicações em neurociência são multiniveladas _____	25
2.1.3	Explicações em neurociência integram múltiplos campos _____	27
2.1.4	CrITÉrios de adequação de uma abordagem explicativa em neurociência _____	27
2.2	Explicação e relevância causal _____	28
2.2.1	O modelo ND frente às normas da explicação em neurociência ___	29
2.2.2	O modelo unificacionista e as explicações em neurociência _____	31
2.2.3	O modelo de Hodgkin e Huxley _____	31
2.3	Relevância causal e manipulação _____	33
2.3.1	Causação como transmissão _____	34
2.3.2	Causação e conexão mecânica _____	36
2.3.3	Causação e manipulação _____	37
2.4	As normas da explicação mecanística _____	38
2.4.1.	Uma alternativa mecanística _____	40
2.4.2.	A descoberta do potencial de ação _____	41
2.4.3.	A caracterização do fenômeno <i>explanandum</i> _____	42
2.4.4.	Componentes _____	43
2.4.5.	Atividades _____	44
2.4.6.	Organização _____	45
2.4.7.	Relevância constitutiva _____	46
2.5	Níveis _____	48
2.5.1	Níveis da ciência _____	49
2.5.2	Níveis da natureza _____	50
2.5.2.1	Níveis da mereologia _____	51
2.5.2.2	Níveis de agregatividade _____	52
2.5.2.3	Níveis de mera continência material/espacial _____	52
2.5.2.4	Níveis de mecanismos _____	52
2.6	Explicação não fundamentalista _____	53

2.6.1	Relevância causal e fazer diferença _____	53
2.6.2	Contrastes e pontos de gatilho _____	54
2.6.3	Poderes causais em mecanismos de alto nível _____	55
2.7	A unidade mosaiforme da neurociência _____	56
2.7.1	A redução na história da neurociência _____	56
2.7.2	As origens da SLP _____	57
2.7.3	A virada mecanística _____	58
2.7.4	O mecanismo como uma proposta de trabalho _____	58
2.7.5	A integração internível e a unidade de mosaico da neurociência _____	58
2.7.6	Redução <i>versus</i> integração de campos no mesmo nível _____	60
Capítulo 3 – críticas ao modelo mecanístico-causal multinívelar _____		63
3.1	Bickle e sua abordagem “implacavelmente redutiva” _____	64
3.2	Schaffner e o papel das teorias, modelos e equações em neurobiologia _____	65
3.3	Weber e as causas sem mecanismos _____	67
3.4	Os mecanismos biológicos de Bechtel _____	71
3.5	Mecanismos e relevância constitutiva _____	73
Conclusão _____		76
Referências Bibliográficas _____		79

Introdução

A filosofia da ciência, enquanto área especializada da filosofia, começou a se consolidar como disciplina na primeira metade do século XIX, atingido o seu ápice no século XX (Abrantes, 2011). Embora questões sobre a explicação científica remontem à antiguidade, esse tema passou a ser reexaminado num novo e importante contexto há pouco mais de meio século, com a elaboração do modelo nomológico-dedutivo, que foi aceito quase unanimemente até poucas décadas atrás. Bem mais recentemente, autores como Wesley Salmon e Philip Kitcher começaram a elaborar abordagens alternativas, com a intenção de resolver alguns dos seus problemas. Isso acabou levantando uma grande quantidade de questionamentos relacionados à explicação em ciência, principalmente indagações a respeito da causalidade e da unidade da ciência.

Num viés contemporâneo, pode-se dizer que a ciência foi tratada como um grupo uniforme, cujo paradigma era a física, então a área mais desenvolvida e capaz de gerar resultados mais robustos. Tanto o modelo nomológico-dedutivo, como a maior parte das alternativas desenvolvidas, incluindo o modelo mecânico-causal de Salmon e o modelo unificacionista de Kitcher, foram elaborados com base nessa área. Outras ciências, como a biologia, a psicologia e a economia, foram ignoradas durante décadas sucessivas.

A partir da segunda década do século XX, a biologia começou a atingir um grau de maturidade e confiabilidade semelhantes aos da física (Weber, 2005). Os primeiros trabalhos em filosofia da biologia, em meados de 1970, concerniam principalmente a biologia evolucionista, o que Weber (2005) atribui a três motivos principais: primeiro, às profundas implicações filosóficas e religiosas que ela traz; segundo, às similaridades estruturais com a física em seus modelos e teorias matematicamente formulados; terceiro, porque biólogos evolucionistas foram pioneiros em escrever textos históricos e filosóficos em biologia.

Talvez por esse motivo, os primeiros textos em filosofia da biologia seguiam o “ideal euclidiano” (Schaffner, 2008), em que a essência de uma teoria é constituída por umas poucas definições e axiomas, estes formulados em símbolos matemáticos e elaborados na forma de teoremas. Com o tempo, porém, percebeu-se que esse tipo de formulação só poderia ser suportado em áreas da física, necessitando de exemplos praticamente atípicos para que fosse adaptada à biologia (Idem).

Logo se tornou evidente que a forma como se produz resultados em biologia é, em muitos aspectos, profundamente diferente da física. Biólogos elaboram problemas distintos dos físicos, conduzem experimentos diferentes e constroem explicações de uma forma muito particular. Por esse motivo, a mera adaptação à biologia de modelos formulados para a física soa como uma

imposição rígida de condições que lhe são alheias e que raramente são cumpridas dentro da prática dos próprios biólogos.

Tornando a questão ainda mais complexa, a biologia em si não é uma ciência uniforme. Existe uma certa independência entre os problemas levantados pela biologia evolucionista e por aquilo que Weber intitula “biologia experimental” e que inclui a genética, a bioquímica, a biologia molecular e a neurobiologia, entre outras.

A neurociência é uma área nova, que começou a ganhar expressão a partir de meados do século XX. A partir de então, começaram a ser publicados estudos em uma velocidade muito grande, com temas que abrangiam desde a descrição do potencial de ação (Hodgkin et al, 1952, a, b, c, d, e), passando pelo aprendizado e memória (Tolman, 1949; Scoville et al, 1959) até a função sináptica (Dale et al, 1936; Katz et al, 1967, a, b; Miledi, 1973).

A Sociedade para a Neurociência (Society for Neuroscience – SfN) foi fundada em 1969, com 500 membros (página da web da SfN). Atualmente, a SfN conta com cerca de quarenta mil membros em cerca de noventa países. Sua missão é:

Aumentar o conhecimento sobre o cérebro e o sistema nervoso, através da união de cientistas de diversas práticas, da facilitação da integração da pesquisa direcionada a todos os níveis de organização biológica e pelo encorajamento de pesquisa translacional e a aplicação de novos conhecimentos científicos para desenvolver melhores tratamentos e curas para doenças (pagina da web da SfN)

Uma busca no portal de periódicos da Capes, feita em 12/03/2013, com o termo “neuroscience” restrito ao campo “assunto” trouxe 6.119 publicações somente no último ano, revelando a dimensão da produção científica nessa área. Nas últimas décadas, a neurociência estabeleceu-se como uma ciência autônoma, com pressupostos e métodos próprios que a separam, em muitos aspectos, das demais áreas da ciência, incluindo a biologia e a psicologia.

Existem algumas particularidades da neurociência que modelos explicativos precisam, necessariamente, levar em conta. Em primeiro lugar, as explicações em neurociência raramente fazem apelo a leis, pelo menos não no sentido forte das leis físicas. No máximo, ela possui generalizações mais ou menos amplas, no sentido usado por Schaffner (1993). Assim, modelos de explicação que fazem apelo a leis gerais, como o modelo nomológico-dedutivo, são inadequados para a neurociência.

Além disso, a neurociência descreve mecanismos. Ou seja, nesse caso, explicar tem a ver com identificar como partes se organizam e se relacionam para gerar o fenômeno que está sendo

observado. Representações matemáticas e simulações, por exemplo, não têm valor como explicação, a não ser que sejam acompanhadas de uma descrição dos mecanismos que elas representam.

Outro aspecto é que a neurociência possui uma orientação fortemente pragmática. Além de satisfazer sua curiosidade intelectual, neurocientistas têm a intenção de oferecer meios para diagnosticar e tratar doenças. Por esse motivo, eles lidam com a causalidade de modo bastante diferente da física: causalidade, aqui, está intimamente relacionada a intervenção, manipulação e controle.

Por fim, a neurociência abrange conhecimentos provenientes de diversas áreas diferentes, incluindo bioquímica, biologia molecular, neurofisiologia, ciências da computação, matemática e psicologia comportamental. Diante dessa heterogeneidade de conhecimentos, é preciso identificar as características que fazem dela uma ciência unificada.

Com esse trabalho, pretendemos examinar um dos modelos mais influentes para abordar as explicações em neurociência: o modelo mecanístico-causal de Carl Craver, conforme apresentado em seu livro *Explaining the Brain* (2007). No primeiro capítulo, faremos uma breve revisão dos principais modelos explicativos clássicos, elaborados para a ciência geral e para as ciências biomédicas. No segundo capítulo, examinaremos detalhadamente o modelo mecanístico-causal. No terceiro capítulo, discutiremos algumas das principais críticas levantadas contra esse modelo. Tais críticas serão retomadas na Conclusão, em contraponto aos seus aspectos positivos, buscando abrir espaço para discussões cada vez mais profundas e esclarecedoras sobre explicações em neurociência.

Capítulo 1 – explicação em ciência geral¹

A busca pelo conhecimento científico remonta à antiguidade, particularmente a Aristóteles, que distinguiu o conhecimento descritivo (conhecimento do *que*) do conhecimento explicativo (conhecimento do *porque*). Entretanto, durante vários séculos, houve muito pouco progresso em se precisar o que, exatamente, constitui uma explicação científica.

Aristóteles já afirmava que argumentos dedutivos podem ser explicações científicas, embora reconhecesse que nem todo tipo de argumento científico constituísse uma explicação científica (Salmon, 2006, p. *xiii*). Contudo, foi após a publicação de “Studies in the Logic of Explanation” por Hempel e Oppenheim, em 1948, que se conseguiu um progresso notável em precisar um tipo de argumento que cumpre esse papel. Mais tarde, isso veio a ser conhecido como o *modelo nomológico-dedutivo*.

Nas décadas que se seguiram, grande parte dos trabalhos filosóficos sobre explicação científica ocorreram como uma resposta mais ou menos direta a esse artigo. Por essa razão, a presente exposição começará por esse ponto.

1.1 O modelo nomológico-dedutivo (ND)

1.1.1 Visão Geral

De acordo com a abordagem nomológico-dedutiva, a explicação de um evento particular é um argumento dedutivo válido cuja conclusão declara que o evento a ser explicado ocorreu. À conclusão, dá-se o nome de *declaração-explanandum*. As premissas são chamadas de *explanans* e devem incluir uma declaração de pelo menos uma lei geral que é essencial à validade do argumento. Um argumento que preenche tais requisitos é chamado de *explicação potencial*. Quando, além disso, as declarações que constituem o explanans são verdadeiras, o argumento pode ser chamado de *explicação verdadeira* ou, simplesmente, *explicação*.

Desde o início, Hempel e Oppenheim reconheceram que nem todas as explicações científicas legítimas pertencem a esse modelo. Algumas são probabilísticas ou estatísticas. Em seu trabalho de 1962 e, posteriormente, de 1965, Hempel e Oppenheim ofereceram a primeira tentativa sistematizada de caracterizar uma explicação probabilística ou estatística. Segundo essa abordagem, há dois tipos de explicação estatística. O primeiro, chamado *estatístico-indutivo*, explica ocorrências

¹ O termo “ciência geral” não se refere a qualquer área específica da ciência. Porém, esses modelos foram desenvolvidos com enfoque direcionado à física. Por esse motivo, quando nos referirmos a ciência geral nesse texto, estaremos falando principalmente da física.

particulares subsumindo-as em leis estatísticas, de forma parecida com o modelo nomológico-dedutivo, exceto em que o argumento é indutivo. O segundo tipo procura explicar, além das ocorrências particulares, regularidades gerais. Nessa concepção, regularidades estatísticas são deduzidas a partir de leis estatísticas mais compreensivas. Esse modelo é chamado de *estatístico-dedutivo* e, para Salmon, pode ser classificado como um subtipo da explicação nomológico-dedutiva.

Hempel (1965) reconhece quatro categorias de explicação científica, que podem ser representadas na tabela abaixo (Salmon, 2006):

<i>Explananda</i>	Fatos Particulares	Regularidades Gerais
Leis		
Leis Universais	Nomológico-Dedutivo (N-D)	Nomológico-Dedutivo (N-D)
Leis Estatísticas	Estatístico-Indutivo (E-I)	Estatístico-Dedutivo (E-I)

Tabela 1. As quatro categorias da explicação científica

Embora na teoria de Hempel (1965) o modelo nomológico-dedutivo possa dar conta da explicação de regularidades gerais e ocorrências particulares, em 1948, Hempel e Oppenheim atêm-se somente às últimas. Para eles, a abordagem explicativa aplicada a regularidades gerais apresenta problemas peculiares que eles se viam incapazes de solucionar.

1.1.2 O modelo

Para se compreender a abordagem ND de Hempel e Oppenheim, é necessário distinguir entre condições gerais de adequabilidade de qualquer instância desse tipo de explicação e a explicação em si (Salmon, 2006). As condições gerais de adequabilidade são as seguintes:

- (1) a explicação deve ser um argumento dedutivo válido;
- (2) o explanans deve conter pelo menos uma lei geral;
- (3) o explanans deve ter conteúdo empírico;
- (4) as enunciados constituindo o explanans devem ser verdadeiros.

A noção geral de uma explicação ND pode ser representada pelo seguinte esquema:

	C1, C2.....Ck	Declarações sobre condições antecedentes
Explanans	L1, L2.....Lk Leis gerais	Descrição do fenômeno empírico a ser explicado
Explanandum	E	

1.1.3 O problema dos enunciados-leis

Um dos principais problemas desse modelo é a caracterização de enunciados-leis. Hempel e Oppenheim (1948), seguindo a tradição de Goodman (1947), começam definindo a noção mais ampla de *enunciados legiformes*, que possuem todas as características de enunciados-leis, exceto a verdade. Assim, um enunciado lei é um enunciado legiforme verdadeiro. Informalmente, enunciados legiformes possuem quatro propriedades:

- (1) possuem forma universal;
- (2) sua abrangência é ilimitada,
- (3) elas não possuem designações de objetos particulares;
- (4) elas contêm predicados puramente qualitativos.

Nesse contexto, é fundamental que se possa distinguir entre leis, que possuem força explicativa, e *generalizações acidentais* que, mesmo quando são verdadeiras, não possuem. Assim, torna-se necessário enumerar quais as características, além da verdade, as leis devem ter. A primeira delas é a generalidade: leis devem poder ser aplicadas universalmente e não devem conter exceções para grupos ou indivíduos. A segunda é a capacidade de suportar contrafactuais, do tipo “o que ocorreria se...”. A terceira característica é a importação modal, ou seja, a capacidade de delimitar o que é necessário, possível ou impossível.

Salmon acrescenta que, mesmo que uma declaração não contenha designações explícitas de objetos particulares, podem envolver referências implícitas a um ou mais particulares, particularmente nos predicados usados. Termos como “lunar” e “solar” são exemplos de termos que se referem a particulares, não sendo, assim, puramente qualitativos.

A proibição de referência a particulares imposta por Hempel e Oppenheim é muito restritiva, como eles próprios admitem. Por esse motivo, eles distinguem entre leis fundamentais e leis derivadas. As restrições acima se referem apenas às leis fundamentais, sendo que as leis derivadas são quaisquer declarações universais que possam ser deduzidas de leis fundamentais.

Ainda assim, para Salmon (1984), Hempel e Oppenheim falharam na distinção entre leis gerais e generalizações acidentais. Universalidade, importação modal e suporte de contrafactuais possuem uma extensão comum. Isso quer dizer que declarações que possuem uma dessas características, necessariamente, possuem as outras duas. Entretanto, ele prossegue, é muito difícil encontrar critérios que permitam distinguir declarações que tenham essas características das que não as tenham. Somente sabendo quais as declarações são legiformes, é possível saber quais possuem as características, e somente sabendo quais as características é possível dizer quais são legiformes. Desse modo, conclui Salmon, a definição é circular².

1.1.4 Contraexemplos famosos ao modelo nomológico-dedutivo

O modelo nomológico-dedutivo foi severamente criticado por teóricos da ciência mais recentes. Um problema óbvio é a questão das relações temporais entre fatos explicativos (o *explanans*) e o fato a ser explicado (o *explanandum*). Embora no esquema acima (p. 3) Hempel e Oppenheim denominem os Cs de “condições antecedentes”, não são feitas restrições temporais quaisquer.

Salmon ilustra essa dificuldade com o exemplo do eclipse. Segundo ele, de acordo com o modelo ND, um eclipse lunar total pode ser explicado a partir das posições relativas da terra, sol e lua, juntamente com as leis mecânicas que governam seus movimentos. No entanto, devido à falta de restrições temporais, o modelo ND permite deduzir a ocorrência do eclipse a partir das posições da Terra, Sol e Lua em um momento *posterior* ao eclipse, o que a maioria dos cientistas consideraria não explicativo.

Outro problema do modelo ND, também mencionado por Salmon, tem a ver com o papel da causalidade nas explicações científicas. A noção senso comum de explicação parece considerar que explicar um certo evento é identificar sua causa. Contudo, embora Hempel e Oppenheim compartilhem dessa ideia, Hempel (1965) defende que nem todas as explicações são causais. Alguns exemplos consagrados na literatura filosófica ilustram essa fragilidade:

(E1) O exemplo do mastro e a sombra:

Considere-se que um mastro de certa altura projeta uma sombra de determinado comprimento quando o sol se situa a uma certa posição. Nessa situação, o comprimento da sua sombra pode ser deduzido a partir da posição do sol, da altura do mastro e da lei da propagação retilínea da luz. Também é possível deduzir a altura do mastro a partir desses parâmetros. As duas

² Embora Salmon desenvolva essa crítica, consideramos que, uma vez que se trata de uma definição, a circularidade é admissível.

situações podem ser consideradas explicações, segundo o modelo ND. Ainda assim, poucos estariam inclinados a aceitar esta última como uma explicação legítima.

(E2) O exemplo do barômetro:

A partir de uma queda abrupta na leitura de um barômetro funcionando corretamente, pode-se inferir que uma tempestade ocorrerá. Ainda assim, a queda no barômetro não explica a tempestade, uma vez que, na realidade, a queda na leitura do barômetro e a ocorrência da tempestade possuem uma causa comum, as condições atmosféricas naquela região.

(E3) O exemplo da Lua e as marés:

Desde muito antes de Newton, é conhecida a correlação entre a posição e a fase da Lua e os movimentos das marés. Mesmo que as relações causais entre esses processos não fossem conhecidas, seria possível construir explicações ND correlacionando a posição e a fase da lua e os movimentos das marés. Ainda assim, aceita-se que só houve uma explicação real desse fato quando Newton descobriu a conexão causal entre esses eventos.

1.1.5 O problema da simetria

Uma das teses mais controversas propostas por Hempel e Oppenheim foi a da simetria entre explicação e previsão. De acordo com ela, o mesmo esquema lógico se adequa à explicação e previsão científicas, sendo a distinção entre elas meramente pragmática. Se o exemplo descrito por E já ocorreu, pode-se perguntar por que ele ocorreu e elaborar uma explicação consistindo na derivação ND de E a partir das condições antecedentes. Se, ao contrário, as mesmas leis e condições antecedentes são dadas antes da ocorrência de E, então o mesmo argumento oferece uma previsão de E.

Como discutido anteriormente, o modelo ND não oferece uma restrição temporal. Por esse motivo, a tese da simetria pode ser formulada de maneira mais abrangente. Nessa concepção, supondo que o explanandum E ocorresse antes das condições C no explanans, condições C poderiam ser usadas para “retrover” E. Seria possível, nesse caso, que E tivesse ocorrido, mas que isso não fosse conhecido. Com o conhecimento das condições C e das leis apropriadas, poder-se-ia descobrir que E realmente ocorreu. Previsão, no sentido mais amplo da tese da simetria, pode ser interpretada como inferência do conhecido para o desconhecido.

Scriven (1959), citado por Salmon (1984) formulou um contraexemplo à tese da simetria:

(E4) A paralisia geral progressiva é uma forma que sífilis terciária, que ocorre somente em indivíduos que tiveram sífilis primária, secundária e latente, sem tratamento com penicilina. Ao se perguntar por que uma pessoa em particular desenvolveu paralisia, a resposta correta seria dizer que ele(a) sofreu de sífilis e não foi tratado(a). Contudo, somente 25% das pessoas com sífilis que não se tratam desenvolvem paralisia. Por esse motivo, a previsão correta para uma pessoa qualquer com sífilis que não faça tratamento é de que ela não vai desenvolver paralisia.

1.1.6 O problema da relevância explicativa:

Em um argumento lógico, a adição de premissas irrelevantes não altera a validade do argumento. Como o modelo ND parte do princípio de que a explicação é um argumento, ele se torna incapaz de distinguir entre condições antecedentes relevantes e irrelevantes. Dois contraexemplos ilustram esse ponto:

(E5) Considere-se que uma amostra de sal de mesa foi colocada na água e se dissolveu. Imagine-se que, antes disso, uma pessoa usando um chapéu engraçado murmurou algumas sílabas sem sentido e abanou as mãos sobre ele, lançando-lhe um feitiço. O modelo ND permite elaborar uma explicação do seguinte tipo:

Todo sal se dissolve em água

Essa amostra é sal

Essa amostra foi enfeitiçada

Essa amostra se dissolveu em água.

Embora esse argumento preencha os requisitos de uma explicação ND, ele não pode ser considerado uma explicação bona fide.

(E6) Um homem chamado John Jones não engravidou no último ano porque estava tomando regularmente as pílulas anticoncepcionais da sua esposa e todo homem que toma pílulas anticoncepcionais previnem-se contra a gravidez. Do mesmo modo que no exemplo anterior, isso também preenche os requisitos do modelo ND, mas não pode ser considerado uma explicação científica.

1.2 Explicação estatística

Em 1962, Rescher fez a primeira defesa expressiva da extensão do conceito de explicação científica além dos limites da explicação dedutiva. Para ele, uma explicação deveria levar em consideração argumentos que tornavam o explanandum meramente provável, e não necessário, como propunha o modelo ND, uma vez que o conceito de explicação defendido por Hempel e Oppenheim era muito restrito e deixava de fora várias considerações explicativas da ciência moderna.

Embora Rescher reconhecesse o caráter indutivo das explicações que propunha, nem ele, nem autores que o seguiram, elaboraram um modelo explícito. Foi o próprio Hempel que, logo depois (1962), notou as profundas dificuldades da explicação estatística, em especial o problema da ambiguidade. No mesmo ano da publicação de Rescher, Hempel ofereceu um modelo explícito das explicações estatísticas. Esse modelo pretende abranger explicações de fatos particulares, caracterizando o que denominou *modelo estatístico-indutivo*. Logo a seguir, introduziu o conceito do seu modelo *estatístico-dedutivo*, com a intenção de também englobar a explicação de leis gerais.

1.2.1 O modelo estatístico-dedutivo

Explicações estatístico-dedutivas (ED) consistem em explicar regularidades estatísticas através da dedução de leis estatísticas mais compreensivas. Por exemplo (Salmon, 1992, p. 23), se uma amostra de madeira possui uma quantidade de carbono-14 (C^{14}) quatro vezes menor do que uma amostra nova de madeira, infere-se que ela tem 11.460 anos. Isso ocorre porque a meia vida do C^{14} é de 5730 anos e, em duas meias vidas, é altamente provável que três quartos dos átomos de C^{14} tenham se degradado, enquanto nas árvores vivas o estoque de C^{14} da atmosfera repõe o estoque da madeira. O argumento explicativo é o seguinte:

Todo átomo de C^{14} que não seja exposto a radiação externa tem uma probabilidade de $\frac{1}{2}$ de se desintegrar em um período de 5730 anos

Em qualquer conjunto de átomos de C^{14} que não seja exposto a radiação externa, aproximadamente três quartos muito provavelmente decairão em 11.460 anos.

O modelo ED é logicamente semelhante ao modelo ND, com a única diferença de que o explanans precisa conter pelo menos uma lei estatística. Contudo, Hempel não desenvolveu esse

modelo após o seu artigo de 1948, deixando-o com lacunas que o comprometem seriamente (Salmon, 1992).

1.2.2 O modelo estatístico-indutivo

A principal preocupação de Hempel era, na realidade, com as explicações EI. Como explica Salmon (1984, p. 53), existe uma maneira natural de estender o tratamento das explicações ND de fatos particulares para abranger também explicações estatísticas de fatos particulares. Nesse caso, o explanans deve incluir pelo menos uma lei estatística. Isso faz com que seja impossível deduzir o explanandum do explanans, uma vez que o argumento se torna indutivo. Ao contrário, um argumento com essa forma mostra que o evento em questão seria previsível caso os fatos explicativos estivessem disponíveis a tempo.

Um exemplo simples desse modelo é (Salmon, 1984 p. 53, 54): Yamamoto sofreu graves danos físicos em agosto de 1945 porque estava a apenas um quilômetro do epicentro da explosão atômica de Hiroshima e qualquer pessoa a essa distância do epicentro de uma explosão atômica daquela magnitude provavelmente sofreria dano físico.

Esse tipo de explicação pode ser descrito como um argumento que torna esperado o evento a ser explicado em virtude de certos fatos explicativos. Entretanto, o modelo estatístico-indutivo possui um problema significativo de ambiguidade. As explicações seguintes ilustram esse fato:

Quase todos os casos de infecção estreptocócica remitem rapidamente após a administração de penicilina

John Jones sofreu uma infecção estreptocócica

John Jones recebeu tratamento com penicilina _____ [r]

John Jones recuperou-se rapidamente.

E:

Quase nenhum caso de infecção estreptocócica penicilina-resistente remite rapidamente após a administração de penicilina

John Jones teve uma infecção estreptocócica penicilina-resistente

John Jones recebeu tratamento com penicilina _____ [r]

John Jones não se recuperou rapidamente.

Salmon (1992) observa que as premissas dos dois argumentos são mutuamente compatíveis, embora as conclusões sejam contraditórias. A origem desse problema reside na natureza dos argumentos estatísticos. Enquanto nos argumentos dedutivos a adição de novas premissas não o invalida, em argumentos indutivos isso pode ocorrer.

1.2.3 Objeções ao modelo estatístico-indutivo

Uma das primeiras críticas formuladas ao modelo estatístico indutivo partiu de Salmon (Salmon, 1984). Para esse autor, o que é crucial em explicações estatísticas não é o quanto o explanans torna o explanandum provável, mas se os fatos citados no explanans fazem diferença na probabilidade de ocorrência do explanandum.

O argumento de Salmon converge com as práticas dos cientistas. Experimentos controlados, no qual um grupo é submetido a determinada terapia enquanto o outro não, são avaliados através da comparação entre os dois grupos ao final do experimento. Por exemplo, se a taxa de remissão de certa doença após um tratamento é alta, mas ainda assim igual à taxa de remissão sem o tratamento, ele não pode ser considerado uma explicação para esse desfecho. Do mesmo modo, se a remissão com o tratamento é improvável, mas significativamente maior do que sem o tratamento, ele deve ser considerado pelo menos parte da explicação. Para ele, o que faz diferença para uma explicação é a relevância estatística, e não a alta probabilidade.

Salmon também argumenta que, ao contrário do que propõe o modelo EI, explicações não são argumentos. Em argumentos, a adição de premissas irrelevantes não compromete a validade do argumento. Embora o requerimento da evidência total para os argumentos indutivos demande a inclusão de todas as premissas relevantes, ele não faz qualquer menção a premissas irrelevantes.

Além disso, Salmon questiona a incapacidade de o modelo EI lidar com eventos de baixa probabilidade. Em processos estocásticos, dois desfechos são possíveis, sendo um provável e um improvável. Ao tratar explicações como argumentos, o modelo EI se revela incapaz de explicar o desfecho improvável. Por exemplo, no caso da sífilis e a paralisia geral progressiva, citado acima, o modelo EI parece explicar a não ocorrência de paralisia, que tem uma probabilidade de 75%, mas não a ocorrência, que tem a probabilidade de 25%.

Por fim, Salmon ressalta a necessidade de imposição de uma assimetria temporal às explicações científicas, o que não é possível quando explicações são tratadas como argumentos. Considera-se que, nas explicações, o explanans ocorre temporalmente antes do explanandum. Como o modelo ND não impõe uma restrição temporal ao antecedente e ao consequente, é possível

elaborar um argumento válido no qual as premissas ocorrem *depois* da conclusão, o que Salmon chama de *inferências*. Contudo, inferências são exatamente opostas a explicações. Enquanto explicações de eventos são feitas em termos de suas causas, inferências são feitas no sentido oposto. A noção de causa possui um valor crucial para as explicações, como defende Salmon (1977), e será melhor desenvolvida adiante.

1.2.4 O modelo da relevância estatística (RE)

Levando em consideração as críticas formuladas ao modelo EI, Salmon (1971) propõe um modelo que leve em conta não a probabilidade de ocorrência do explanans, mas a relevância estatística de um fator para a ocorrência do explanans. Tome-se o seguinte exemplo:

(E7) Suponha-se que Bruce Brown possua um sintoma neurótico importante. Ele é submetido a psicoterapia e seu sintoma remite. É possível dizer que a psicoterapia constitui uma explicação da sua melhora?

Para responder a essa questão, Salmon estabelece alguns pontos: Bruce Brown é um membro da classe de pessoas que possuem um sintoma neurótico do tipo N. Nessa classe, independente do tratamento a que é submetida, a pessoa tem uma probabilidade R de melhorar. A isso, Salmon dá o nome de $\Pr(R/N)$. Existe ainda a probabilidade de recuperação das pessoas com os sintomas que façam psicoterapia, que ele chama de P, que pode ser simbolizada como $\Pr(R/N.P)$. Assim, se $\Pr(R/N.P) > \Pr(R/N)$, então a psicoterapia é positivamente relevante para a recuperação. Se $\Pr(R/N.P) < \Pr(R/N)$, então a psicoterapia é negativamente relevante para a recuperação. Se $\Pr(R/N.P) = \Pr(R/N)$, então a psicoterapia é irrelevante para a recuperação. Somente no primeiro caso, é possível dizer que a psicoterapia explica a recuperação.

Uma das coisas mais notáveis desse modelo é que ele não constitui um argumento. Não há, nesse caso, premissas ou conclusões, apenas explanans e explanandum. Contudo, Salmon acrescenta que, embora o modelo RE contenha vantagens em relação ao EI, ele ainda apresenta uma deficiência crucial: o foco na relevância estatística, e não na relevância causal. Segundo esse autor, a relevância causal é que possui poder explicativo genuíno. A relevância estatística, para ele, é apenas uma evidência da relevância causal.

1.3 Explicação Causal

Nos seus trabalhos mais recentes, Salmon abandonou a tentativa de caracterizar explicação e causalção em termos puramente estatísticos. Ele desenvolveu um modelo, que chamou de modelo mecânico-causal (MC), que pretendia captar o “algo mais” envolvido em relações causais e explicativas. Esse modelo guarda muitas semelhanças com as chamadas teorias do processo, como a defendida por Dowe (2000) (Woodward, 2010).

Ao chamar atenção para a importância da causalidade nas explicações científicas, Salmon reconhece a necessidade de retomar as questões levantadas por Hume sobre a natureza das relações causais: se existe algo que distingue eventos meramente contíguos de eventos causais, isso deve ser algum tipo de conexão física entre eles. Embora Hume não tenha encontrado nenhum elemento adicional que caracterize a causalidade, além de prioridade, contiguidade e conjunção constante, Salmon (1977) segue um caminho diferente.

Para ele, a melhor maneira de esclarecer essa diferença é distinguir *processos causais* de *pseudo-processos*. Um processo pode ser compreendido da seguinte forma: considerando que um evento é algo que ocorre em uma região do espaço-tempo mais ou menos restrita, a depender do contexto, um processo é algo que, naquele contexto, possui uma duração maior do que o evento. O que diferencia processos causais genuínos de pseudo-processos é a sua capacidade de transmitir marcas de forma temporoespacialmente contínua. Aqui, marca pode ser uma modificação local na estrutura do processo, por exemplo, um arranhão na superfície de uma bola de baseball ou um aumento na quantidade de energia de um átomo.

Também é importante a noção de transmissão. Considera-se que um processo é capaz de transmitir uma marca se, uma vez que a marca é introduzida em uma localização espaço-temporal, ela persiste em outras localizações espaço-temporais mesmo na ausência de outras interações posteriores.

É importante ressaltar que, como o próprio Salmon (2006) admite, a habilidade de transmitir marca é uma noção contrafactual. Em primeiro lugar, um processo pode ser considerado um processo causal mesmo que ele de fato não transmita qualquer marca, desde que seja verdadeiro que, se fosse adequadamente marcado, ele transmitiria a marca. Além disso, a própria noção de marca envolve um contraste contrafactual, ou seja, um contraste entre como um processo se comporta quando marcado e como ele se comportaria se permanecesse sem marcação.

Outra noção importante para o modelo de Salmon é a de *interação causal*. Uma interação causal envolve uma intersecção espaço-temporal entre dois processos causais que modifica a estrutura de ambos, que adquirem características que não teriam na ausência da interação.

De acordo com o modelo MC, uma explicação de um evento E traça os processos e interações causais que levam a E, ou pelo menos uma parte deles, além de descrever os processos e interações que constituem o evento. Nesse sentido, a explicação pode ser vista como aquilo que mostra como E se encaixa em um nexos causal.

Para Woodward (2010), embora a noção de “encaixe em um nexos causal” não seja muito precisa, a aplicação da ideia intuitiva a casos individuais é clara. Por exemplo, suponha-se que uma bola de bilhar, posta em movimento por um taco, atinja uma bola oito parada, resultando no movimento da bola oito e na mudança de direção da primeira bola. Além disso, o impacto do movimento transmite giz azul à primeira bola que, por sua vez, transfere-o para a bola oito no impacto. O taco, a primeira bola e a bola oito são processos causais, o que pode ser ilustrado pela transmissão da marca de giz, e a colisão do taco com a primeira bola e desta com a bola oito são interações causais. Para Woodward, a ideia de Salmon é que citar tais fatos sobre processos e interações explica os movimentos das bolas após a colisão. Ao contrário, se uma das bolas exibe uma sombra sobre a outra, isso será causalmente e explicativamente irrelevante, uma vez que a sombra é um pseudo-processo.

1.3.1 A concepção ôntica da explicação

Coffa (1977) foi o primeiro filósofo a elaborar uma concepção ôntica sobre a explicação científica, contrastando com a concepção epistêmica da expectabilidade científica. (Salmon, 2006), De acordo com essa visão, os eventos a serem explicados ocorrem em um mundo repleto de regularidades que são causais, leis ou ambas. Tais regularidades podem ser determinísticas ou estatísticas. Em ambos os casos, explicar o evento consiste em situá-los nas estruturas existentes no mundo. A busca por explicações decorre do desconhecimento de alguma parte dessas estruturas; a explicação é feita oferecendo informações sobre tal estrutura e encaixando nela o evento explanandum.

Salmon (1978) defende que, para oferecer uma explicação ôntica sobre um evento, são necessários conhecimentos causais: o reconhecimento da diferença de leis causais e leis não causais, a diferença entre processos causais e pseudo-processos e a diferença entre interações causais e meras coincidências espaço-temporais. Desse modo, de acordo com a concepção ôntica, conhecimento explicativo é o conhecimento dos mecanismos causais e, possivelmente, mecanismos de outros tipos.

Segundo a concepção ôntica da explicação, existe uma grande diferença entre explicação e previsão. Embora a subsunção a uma lei seja suficiente para elaborar previsões, explicações só são

possíveis com conhecimentos sobre a relação causal existente entre o explanans e o explanandum.

Além disso, a concepção ôntica admite que explicações não são epistemicamente relativizadas, não possuem componentes psicológicos, nem dimensões pragmáticas. Explicações refletem fatos objetivos do mundo.

1.3.2 O modelo mecânico-causal e a relevância explicativa

Exemplos como as colisões nas quais existe ação por contato são os paradigmas do modelo MC (Woodward, 2010). Contudo, esse modelo possui uma falha importante. Como observa Hitchcock (1995), explicações científicas sobre os movimentos das bolas são feitas através da derivação daquele movimento a partir das suas massas e velocidades antes da colisão, da assunção de que a colisão é perfeitamente elástica e da lei de conservação do momento linear. O modelo MC, por sua vez, considera que a transmissão de marca (como um arranhão em uma bola de baseball) é o que diferencia um processo de um pseudo-processo. Ainda assim, estamos inclinados a considerar que as informações relevantes para a explicação são as primeiras, e não a transmissão da marca.

Entretanto, o modelo MC tem dificuldade de captar essa diferença. Isso ocorre porque, para expressar esse tipo de julgamento, é necessário abordar relações entre propriedades e magnitudes e não está claro como expressar tais relações em termos de fatos sobre processos e interações causais. Tanto o momento linear como o arranhão são marcas transmitidas pelo processo causal temporoespacialmente contínuo que consiste no movimento do taco e da bola. As duas marcas são transmitidas através de uma interação entre os dois objetos. Entretanto, não há nada na noção de Salmon de transmissão de marca ou do processo causal que permita distinguir entre o momento, relevante explicativamente, e o arranhão, explicativamente irrelevante.

Esse tipo de problema também pode ser observado no exemplo da pílula anticoncepcional, usado por Salmon para mostrar a falha do modelo ND em captar noções de relevância explicativa. Para ele, processos causais temporoespacialmente contínuos que transmitem marcas estão presentes quando Mr. Jones ingere a pílula anticoncepcional (ela se dissolve, é absorvida, metabolizada, etc). Do mesmo modo, quando Mrs. Jones ingere a pílula anticoncepcional, processos causais temporoespacialmente contínuos que transmitem marcas estão presentes, embora possam ser diferentes dos primeiros. Nesse sentido, a relevância ou irrelevância das pílulas anticoncepcionais prevenirem uma gestação não pode ser captada simplesmente ao indagar se os processos que levam a esse desfecho são processos causais, no sentido de Salmon.

De modo geral, essa questão pode ser colocada da seguinte forma: considerando que P é um processo causal, M é uma marca transmitida, E o evento a ser explicado e R uma propriedade

qualquer, as características de P em virtude das quais ele se qualifica como processo causal (a capacidade de transmitir a marca M) podem não ser as mesmas características de P que são causalmente ou explicativamente relevantes para o desfecho de E que se quer explicar. M pode ser irrelevante para E , sendo outra propriedade R aquela relevante para a explicação. Por esse motivo, esse modelo não oferece meios de distinguir características e propriedades de um processo causal que sejam causalmente ou explicativamente relevantes para um desfecho daquelas que não o são.

1.3.3 O modelo mecânico-causal e os sistemas complexos

Para Woodward (2010), o modelo MC exibe dificuldades ao ser aplicado a sistemas que se distanciam dos casos paradigmáticos apresentados por Salmon. Um exemplo são as teorias, que envolvem ação à distância, como a da gravitação Newtoniana. Além disso, Woodward cita vários exemplos que envolvem interações causais sem processos têmporo-causalmente contínuos ou transferência de energia e momento da causa para o efeito, como causação por omissão e causação por prevenção, que serão discutidos no capítulo 2.

Woodward acrescenta que o modelo MC também tem dificuldade de lidar com sistemas complexos, ou de nível alto, como sistemas biológicos, psicológicos, econômicos e até mesmo físicos, que não citam explicitamente processos causais têmporo-espacialmente contínuos que envolvem transferência de energia e momento, ainda que tais processos ajam em um nível subjacente. Por exemplo, a expansão isotérmica de um gás não é explicada através da identificação de cada um dos processos causais de cada uma das 6×10^{23} moléculas e as interações sucessivas entre elas, mas sim com base nas generalizações da termodinâmica.

1.3.4 O modelo das quantidades conservadas (QC)

Em 1994, Salmon elaborou uma teoria da causação que buscava evitar qualquer tipo de apelo a contrafactuais. Nesse modelo, Salmon definiu um processo causal como um processo que transmite uma porção não nula de *quantidade conservada* em cada momento de sua história. Quantidades conservadas são, segundo ele, quantidades assim caracterizadas na física, como momento linear, momento angular, carga, etc. Um processo transmite quantidade conservada de A para B se possui aquela quantidade em cada estágio sem qualquer interação que envolva uma troca de quantidade no intervalo semi-aberto (A,B].

A dificuldade mais fundamental desse modelo é que ela também é inadequada para lidar com o problema da relevância explicativa e da relevância causal. Em 1997, Salmon tentou resolvê-

la. Ele sugeriu que essas noções podem ser captadas ao apelar para a noção de processo causal juntamente com informação sobre relações de relevância estatística. Contudo, para Woodward, essa ideia não está desenvolvida com detalhes no trabalho de Salmon, que não deixa claro como ela pode ser posta em funcionamento. Segundo ele, enquanto for possível que diferentes alegações causais impliquem nos mesmos fatos sobre relações de relevância causal, essa nova proposta está fadada a fracassar também.

1.4 Explicação como unificação

A noção de unificação foi inicialmente proposta por Friedman (1974) e, posteriormente, desenvolvida por Kitcher, 1989. De acordo com eles, o conhecimento científico fornece explicações ao reduzir o número de fatos últimos que se precisa levar em conta (Friedman, 1974, 18). Isso quer dizer que explicar consiste em unificar diversas crenças sob um conjunto simples de argumentos (Craver, 2007). Nas palavras de Kitcher:

Compreender um fenômeno não é simplesmente uma questão de reduzir as “incompreensibilidades fundamentais”, mas de ver conexões, configurações comuns, naquilo que inicialmente pareciam ser situações diferentes. Aqui, a mudança na concepção de pares premissa-conclusão para derivações se mostra vital. *A ciência aumenta o nosso conhecimento sobre a natureza ao mostrar-nos como derivar descrições de muitos fenômenos, usando o os mesmos padrões de derivação sucessivamente e, ao demonstrá-lo, ela nos ensina como reduzir o número de fatos que precisamos aceitar como últimos (ou brutos).* (Kitcher, 1989, p. 501, grifo do autor).

A ideia central é que a unificação da ciência é feita através da restrição cada vez maior dos arranjos de argumentos³ necessários para explicar uma quantidade cada vez maior de fatos. Talvez o mais conhecido exemplo seja a unificação das leis de Kepler e Galileu pelas leis de Newton. Antes, elas eram consideradas separadamente. Contudo, as leis de Newton implicam nelas e diminuem a quantidade de regularidades independentes que são aceitas, unificando-as.

Os detalhes da formalização do modelo estão descritos em Kitcher, 1989, p. 431-448. De

3 No vocabulário de Kitcher, *frases esquemáticas* são frases nas quais termos não-lógicos foram substituídos por letras. *Instruções de preenchimento* são direções que especificam como inserir as letras nos lugares dos termos não-lógicos. *Argumentos esquemáticos* são sequências de declarações esquemáticas. *Classificações* são aquilo que descreve quais declarações nos argumentos são premissas e conclusões e quais as regras de inferência são usadas. *Arranjo de argumento* é uma tripla ordenada que consiste de um argumento esquemático, um conjunto de instruções de preenchimento e a classificação do argumento esquemático.

modo geral, considere-se que o K seja o conjunto de conhecimentos aceitáveis em determinados tempos. A sistematização de K é feita ao desenvolver um ou mais arranjos de argumentos que podem ser usados para derivar alguns membros de K de outros, com o menor conjunto de arranjos de argumentos possível para derivar todos os membros de K. Kitcher chama esse conjunto mínimo de estoque explicativo (E(K)).

1.4.1 O problema das assimetrias

O modelo unificacionista enfrenta dificuldades para lidar com alguns aspectos da explicação científica, como as assimetrias. O exemplo do mastro e da sombra ilustra como isso ocorre. Nossas práticas explicativas, P, comprometem-se com a ideia de que derivar o tamanho de um mastro a partir da sua sombra não é explicativo. Kitcher compara P com uma sistematização alternativa na qual essa derivação é considerada explicativa. Segundo ele, P inclui o uso de um arranjo de explicação único que ele chama de origem e desenvolvimento (OD), segundo o qual os tamanhos dos objetos são traçados até as condições sob as quais ele se originou e as modificações aos quais ele foi submetido. Ao adicionar a P um arranjo adicional S que permita a derivação das dimensões dos objetos a partir das suas sombras, haverá um aumento do número de padrões de argumentos sem aumentar o número de conclusões possíveis. Por outro lado, se OD for retirado de P e substituído por S, o número de padrões permanece o mesmo, mas o número de conclusões possíveis diminui consideravelmente, uma vez que vários objetos não possuem sombras das quais derivar suas dimensões. Por esse motivo, o sistema P constitui uma explicação, ao contrário dos demais.

Kitcher trata outros problemas de assimetria da mesma forma. Qualquer explicação do porquê uma amostra de sal se dissolve em água que apele para o fato de ele ter sido enfeitado (H) é defeituosa, principalmente se comparada à explicação padrão que apela para generalizações de que todo sal se dissolve em água (D). Isso ocorre porque qualquer derivação que apele para H instancia um arranjo de argumento que pertence a uma totalidade de arranjos que é menos unificadora que a totalidade contendo a derivação que apela para D. As consequências de um estoque explicativo que contenha H, mas não D, são mais restritas do que as de um estoque que contenha D, mas não H, uma vez que a última é incapaz de explicar a dissolução de sal não enfeitado em água. Além disso, a adição de H a um estoque explicativo que contenha D aumenta o número de arranjos sem, contudo, compensar com mais derivações.

Para Woodward (2010), o modelo unificacionista possui um apelo intuitivo importante, mas enfrenta uma série de limitações. Uma delas é a dependência de verdades contingentes, por exemplo, de que alguns objetos não projetam sombras. Isso se torna mais claro usando-se outro

exemplo, proposto por Barnes (1992), no qual existem tantas derivações dos efeitos para as causas como das causas para os efeitos. Considere-se uma teoria como a mecânica newtoniana aplicada a um sistema fechado. Derivações dos movimentos dos planetas em um tempo futuro a partir de informações sobre sua posição presente são chamadas preditivas. Derivações no sentido oposto, ou seja, de um tempo futuro para o tempo presente, são chamadas retroditivas. Nesse caso, existem tantas derivações preditivas como retroditivas e os padrões de derivação de ambas possuem o mesmo número de arranjos. Contudo, considera-se que as derivações preditivas, mas não as retroditivas, são explicativas.

1.4.2 Unificação não explicativa

Woodward (2010) e Craver (2007) ressaltam que, em ciência, unificação constitui uma noção heterogênea, havendo vários tipos de unificação que não são explicativas. Por exemplo, a taxonomia de Lineu sistematiza e unifica o universo biológico, mas não o explica (Craver, 2007). Também constitui uma unificação a criação de um formalismo matemático que possa ser aplicado a diferentes fenômenos, como os sistemas de equação formulados por Lagrange e Hamilton.

Para Craver (2007), o que distingue esses exemplos das unificações explicativas é, precisamente, o fato de as últimas citarem relações causais, como no caso da física newtoniana citado acima. O modelo de Kitcher, contudo, não se compromete com a noção de causalidade. Para distinguir explicações etiológicas (como a explicação do tamanho da sombra a partir do mastro) de meras inferências, ele recorre principalmente a fatos contingentes sobre o mundo.

1.5 Explicações em medicina e ciências biomédicas

Até o final da década de 1980, praticamente toda a literatura sobre explicação em ciência estava centrada nas ciências básicas, especificamente a física. Somente a partir da década de oitenta começou a haver um crescimento expressivo do interesse por explicações em biologia e ciências biomédicas, culminando com a publicação de “Discovery and Explanation in Biology and Medicine” (Schaffner, 1993). Depois disso, diversos trabalhos subsequentes foram publicados, centrados principalmente em explicações em evolução (Sober, 2008), em biologia molecular (Weber, 2005) e em neurociências (Machamer, 2000). Nesse item, faremos um rápido apanhado sobre explicações em biologia em medicina, centrado principalmente na abordagem desenvolvida por Schaffner (1993). Interessantemente, o exemplo usado por Schaffner para desenvolver sua

abordagem é um fenômeno neurobiológico: o mecanismo de formação da memória na *Aplysia*⁴.

De acordo com Schaffner, uma das principais características que distingue as explicações em biologia daquelas em física é a ausência de leis propriamente ditas. Ao invés disso, há generalizações mais ou menos amplas, por exemplo, o mecanismo de ação das proteínas quinase através da fosforilação de moléculas. Essas generalizações tipicamente estão implícitas, podendo ser abordadas explicitamente em capítulos introdutórios de neurociências. Por sua vez, essas generalizações são derivadas de um amplo espectro de modelos, como o modelo de síntese de proteínas e os modelos bioquímicos, por exemplo. Mesmo quando essas generalizações são deixadas explícitas, elas precisam ser complementadas com detalhes específicos sobre o sistema a ser explicado.

Nesse tipo de explicação, existem sequências causais internivelares, nas quais a explicação para um fenômeno é dada em termos moleculares. Nesse sentido, um fenômeno como o aprendizado é reinterpretado em um vocabulário diferente, pertencente a um nível mais baixo. No caso, o aprendizado da *Aplysia* é reinterpretado como uma instância de um fenômeno neuronal, onde coexistem partes de vários níveis (bioquímico, celular, orgânico, etc). É interessante observar que, à medida que se descende nos níveis (saindo do nível celular para o bioquímico, por exemplo), existe um aumento do grau de detalhamento apreendido. No entanto, não existe uma sequência causal que sai dos níveis inferiores para os níveis superiores, como os fenômenos bioquímicos causando os fenômenos celulares. Não há, necessariamente, uma relação causa e efeito entre as partes e o todo que elas compõem. Sequências causais podem ser encontradas dentro e entre quaisquer níveis de agregação.

Nas explicações biomédicas, não há, portanto, generalizações fortes como as leis de Newton, mas sim “uma rede de *sistemas* intrincados a qual se aplicam *generalizações causais* amplas e estreitas que são moldadas não em terminologia bioquímica pura, mas numa terminologia que é caracteristicamente *internível e intercampo*” (Schaffner, 1993, p. 285). Isso é o que ele chama de *modelo parcial de redução com generalizações amplamente implícitas*. Ainda que não seja uma redução completa, por exemplo, até o nível bioquímico, existe uma redução para um modelo de nível mais baixo, tipicamente internivelar. Além disso, os próprios modelos podem não ser robustos para esse ou outros organismos, como ocorre em modelos da física, mas possuem um domínio estreito de aplicação.

A explicação consiste, então, em uma rede de generalizações causais de abrangência maior ou menor e que é expressa por um sistema *idealizado*. Nesse sentido, explicações são aspectos

4 A *Aplysia* é um molusco marinho frequentemente usado como um modelo animal para o estudo de certos eventos celulares do sistema nervoso. O mecanismo de formação da memória da *Aplysia* será discutido com detalhes adiante, quando falaremos especificamente de explicações em neurociência.

parciais de um conjunto mais amplo, que são escolhidos, frequentemente, com base em interesses pragmáticos.

Além disso, Schaffner (1993) aponta que, nas ciências biomédicas de modo geral, as explicações formuladas assemelham-se àquelas propostas pelo modelo mecânico-causal de Salmon (1984). Todavia, existe uma diversidade muito grande de mecanismos, alguns com aplicação muito estreita, com variações até mesmo dentro do próprio organismo, outros com aplicação quase universal. Ainda assim, ele acredita que isso não diminui a força dos mecanismos e generalizações como instrumentos de explicação. A redução e a explicação em ciências biomédicas ocorrem quando “(comparativamente) poucos mecanismos implícitos dão conta de uma (comparativamente) grande diversidade de fenômenos através da apresentação de trajetos causais (complexos)”. (Schaffner, 1992, p. 339).

Capítulo 2

Para Carl Craver (2007), teorias neurocientíficas aceitáveis devem ter três principais características: descrever mecanismos, abranger múltiplos níveis e abranger múltiplos campos. Para o autor, ao avaliar a adequação de teorias neurocientíficas, deve-se ter dois objetivos em mente. O primeiro objetivo é descritivo. Nesse sentido, é necessário identificar as características de explicações em neurociência contemporânea, assim como os critérios com que os neurocientistas as avaliam. Para isso, ele examina casos clássicos da história recente da neurociência, particularmente o potencial de ação e a potenciação de longa duração.

Uma abordagem explicativa em neurociência precisa refletir a prática dos cientistas, mas também orientar a forma como novas explicações serão elaboradas. Neurocientistas cometem erros e discordam entre si sobre a validade de uma explicação proposta. Ademais, os próprios critérios para avaliar a adequação de explicações mudam ao longo da história. Por isso, além de descrever a forma como são conduzidas pesquisas em neurociências, é necessário oferecer uma maneira de distinguir entre boas e más explicações. Isso é o que Craver chama de projeto normativo. Assim, não é suficiente enunciar as normas de explicação em neurociência a partir da maneira como os cientistas as propõem.

A relação entre o projeto descritivo e o projeto normativo é complexa. Faz-se necessário examinar casos incontestados de explicações bem e mal sucedidas, apreendendo o que existe de comum entre boas explicações e entre explicações ruins e o que diferencia boas e más explicações. A partir disso, é possível elaborar normas de orientem a prática científica.

Além disso, a busca de critérios normativos deve levar em conta o caráter pragmático da neurociência. A atividade neurocientífica é conduzida não apenas pela curiosidade sobre o funcionamento do cérebro, mas também com a intenção de intervir, de alguma forma, no seu funcionamento. Neurocientistas procuram oferecer meios de diagnosticar e tratar doenças e danos cerebrais, melhorar a função do cérebro ou prevenir o declínio do seu funcionamento. Uma vez que esses objetivos são indissociáveis da prática neurocientífica, eles precisam ser levados em conta na elaboração de métodos para avaliar suas explicações.

É preciso deixar claro, ainda, que termo “explicação” é ambíguo. Em determinados contextos, o termo é usado para se referir a textos, como descrições, modelos ou representações de qualquer tipo. Nesse caso, “explicação” tem o sentido de “texto explicativo”. Outras vezes, ele é usado para se referir a uma porção objetiva da estrutura causal do mundo, à sequência de fatores que dá origem ao fenômeno, ou seja, ao próprio mecanismo, o que Craver chama de “explicações objetivas”. Explicações objetivas seriam, então, coisas corpóreas, fatos, e não representações. São

coisas que podem ser descobertas e descritas. Explicações objetivas não são certas ou erradas, nem boas ou más. Elas simplesmente são todas e apenas as porções da estrutura causal do mundo relevantes para a ocorrência do fenômeno. Por outro lado, textos explicativos podem ser melhores ou piores, mais ou menos completos. Bons textos explicativos são aqueles que se aproximam mais de descrever a estrutura causal do fenômeno.

Nos próximos itens, será exposta detalhadamente a abordagem explicativa proposta por Craver (2007), preparando para uma discussão crítica no capítulo seguinte.

2.1 Visão geral

2.1.1 Explicações em neurociência descrevem mecanismos

Muitas vezes, ao elaborar explicações sobre o cérebro, os cientistas recorrem aos mecanismos, como ilustram alguns exemplos: *Compartmentalization within neurites: Its mechanisms and implications* (Katsuki et al. 2011), *Retinoic acid receptor signaling regulates choroid fissure closure through independent mechanisms in the ventral optic cup and periocular mesenchyme* (Lupo et al. 2011), *Anabolic androgenic steroid abuse: multiple mechanisms of regulation of gabaergic synapses in neuroendocrine control regions of the rodent forebrain*. (Oberlander et al. 2011). Embora outros termos, como “sistemas”, “cascatas”, “modulação” e “trajeto” sejam comumente usados, eles mantêm o mesmo sentido de “mecanismo”.

“Mecanismo” é um termo amplamente usado ao longo da história, com diferentes sentidos, muitos dos quais sem aplicação na biologia. Discute-se, por exemplo, se mecanismos são abstratos ou concretos, sobre a relação entre mecanismo e teleologia e se a doutrina de mecanismos pode ser tida como um método científico ou uma tese metafísica (Craver, 2007).

No sentido usado por Craver, mecanismos consistem em um conjunto de entidades e atividades organizadas de tal forma que exibem o fenômeno a ser explicado. Um exemplo é o mecanismo de liberação de neurotransmissores por um neurônio, que pode ser descrito da seguinte forma: o potencial de ação despolariza o terminal do axônio e abre canais de Ca^{2+} voltagem-dependentes na membrana. À medida que a concentração de Ca^{2+} aumenta, ele se liga a uma proteína chamada calmodulina, que muda de conformação, formando o complexo cálcio-calmodulina. Esse complexo ativa uma proteína cinase dependente de cálcio-calmodulina que, por sua vez, fosforila a sinapsina, a qual libera a vesícula do citoesqueleto. Quando isso ocorre, a vesícula é guiada para os sítios de liberação por duas proteínas, Rab₃A e Rab₃C. Nesse ponto, proteínas v-SNARES incorporadas na membrana vesicular ligam-se a t-SNARES incorporadas na

membrana axônica terminal, aproximando a vesícula do terminal axônico. Então, o influxo de Ca^{2+} leva à abertura de um poro de fusão que liga a membrana à fenda sináptica, liberando o neurotransmissor.

Observe-se que, nessa explicação, procura-se incluir todos e apenas os componentes e suas ações que fazem diferença para a ocorrência do fenômeno, que são as entidades e as atividades do mecanismo. As entidades são as partes ou componentes dos mecanismos. Elas tem propriedades que lhes permitem várias atividades. Têm localizações, tamanhos, estruturas e orientações. São coisas que têm massa, carga e transmitem momento. Elas podem agir de várias formas: podem ligar-se a outras entidades, abrir-se, fechar-se e difundir-se. Nesse caso, o axônio, os canais de cálcio, a calmodulina, o íon Ca^{2+} , as proteínas Rab₃A, Rab₃C, v-SNARES e c-SNARES, por exemplo, são atividades do mecanismo de liberação do neurotransmissor.

Por sua vez, atividade, de acordo com Craver, é um termo usado para descrever comportamentos produtivos (como abertura), interações causais (como atração), omissão (como nos casos de inibição), prevenção (como no caso do bloqueio). O termo “produtivo” é usado no sentido de que as atividades não são meras correlações, sequências temporais e, principalmente, que podem ser potencialmente exploradas para manipulação e controle. Além disso, as entidades e atividades nos mecanismos são organizadas espacialmente, temporalmente, causalmente e hierarquicamente.

Assim, mecanismos são entidades e atividades organizados de tal maneira que resultam no fenômeno a ser explicado. Esquemáticamente, pode-se chamar o fenômeno, ou seja, a propriedade ou comportamento a ser explicado, de ψ . O mecanismo como um todo, pode ser chamado de S, as atividades, de ϕ e as entidades, de X. Colocando-se dessa forma, pode-se dizer que a ψ -ação de S é explicada pela organização das entidades $\{X_1, X_2, \dots, X_m\}$ e atividades $\{\phi_1, \phi_2, \dots, \phi_n\}$.

2.1.2 Explicações em neurociência são multiniveladas

As explicações em neurociência tratam dos comportamentos dos organismos, das funções de processamento dos sistemas cerebrais, das propriedades representacionais e computacionais das regiões cerebrais, das propriedades eletrofisiológicas dos neurônios e das mudanças das estruturas e conformações das moléculas.

Dizer que explicações em neurociência são multiniveladas significa que é necessário usar elementos de vários níveis para explicar os fenômenos. Craver, citando Stricker e Verbalis, oferece como exemplo um esboço explicativo do mecanismo da sede para a manutenção do equilíbrio hidroeletrólítico:

Um aumento da osmolalidade plasmática, como quando alguém come comida salgada ou após a água corporal evaporar sem ser repostada, estimula a liberação de vasopressina (da pituitária), aumentando a conservação de água e excreção de solutos na urina. Isso é acompanhado de sede aumentada, com o resultado de fazer a osmolalidade plasmática mais diluída através do consumo de água.

Aqui, a explicação é composta por fenômenos como comportamentos (beber água), necessidades fisiológicas (sede), funcionamento de órgãos (conservação da urina nos rins), fluxo de moléculas (liberação de vasopressina pela pituitária) e movimentos iônicos (concentração de sal no sangue). É interessante observar que, ao contrário do que reza a tradição em ciências tradicionais, como a física e a química, aqui não há necessidade de teorias completas sobre cada nível (átomos, moléculas, células, órgãos e organismos) para dar uma explicação completa do fenômeno. As explicações em neurociência são fragmentárias e levam em conta apenas os itens relevantes para cada fenômeno ou cada aspecto do fenômeno.

Para Craver, a multinivelaridade é uma característica imprescindível para boas explicações em neurociência. Não existe um nível básico no qual a explicação ocorre, seja o nível dos neurônios, moléculas ou átomos. Níveis mais altos não possuem apenas valor heurístico, mas são necessários para a explicação adequada de vários fenômenos do sistema nervoso. Essa posição é radicalmente oposta ao chamado fundamentalismo⁵ radical, segundo o qual a única explicação real se situa em um nível fundamental.

O fundamentalismo é amplamente aceito na neurociência moderna, mas carece de evidências empíricas. Craver afirma que, muitas vezes, a defesa de uma posição fundamentalista decorre de algum “orgulho disciplinar” motivado pela crença de alguns cientistas de que sabem mais sobre o mundo do que os de outras áreas, ou de que suas técnicas funcionam melhor que as outras.

Todavia, outras vezes ela decorre de uma posição metafísica, como a defendida por Jaegwon Kim. Segundo ele, tudo possui uma causa completa no nível fundamental (princípio da completude causal da física) e nada possui mais do que uma causa completa (princípio da não-superdeterminação). Assim, não há espaço para o poder causal de objetos situados em um nível não fundamental. Essa posição é problemática para a neurociência, uma vez que neurocientistas não trabalham em um nível fundamental

Há também argumentos históricos a favor do fundamentalismo. A ciência teria uma tendência a explicações em termos de unidades cada vez mais fundamentais. Esse seria o caso de

⁵ Craver usa o termo *fundamentalism* (que traduzimos como *fundamentalismo*) referindo-se à noção de que boas explicações só podem ser formuladas no nível mais fundamental. Como ele aponta, esse termo é sinônimo de reducionismo e fisicalismo Craver, 2007, p. 11, n.r. 13).

explicações de fenômenos sociais em termos de fenômenos psicológicos e desses últimos em termos de neurônios, por exemplo. Por fim, Craver acredita que o fundamentalismo poderia ser uma reação contra o “medo do escuro”, referindo-se à noção infundada de que entidades de nível superior são algo como enteléquias fantasmagóricas, almas ou propriedades emergentes. Esse tema será retomado adiante.

2.1.3 Explicações em neurociência integram múltiplos campos

A Society for Neuroscience oferece a seguinte definição de neurociência: “Neurociência, o estudo do sistema nervoso central, abrange a compreensão do pensamento, da emoção e do comportamento humano”. Segundo essa sociedade, a neurociência usa aspectos da anatomia, psicologia comportamental, biofísica, psicologia do desenvolvimento, ciência da computação, biologia molecular, endocrinologia, etologia, imunologia, neurologia, neurofisiologia, matemática, farmacologia, física, fisiologia e psiquiatria para explicar como o sistema nervoso se desenvolve e funciona normalmente e o que acontece de errado nos transtornos neurológicos.

Tradicionalmente, aceita-se que a unidade da neurociência é obtida através de uma cadeia de explicações redutivas que ligam fenômenos de nível alto a fenômenos de nível baixo. A redução é obtida através da identificação dos tipos de termos nas teorias de nível alto com aqueles das teorias de nível baixo. Muitos filósofos seguiram essa ideia, proposta por Oppenheim e Putnam (1958).

Contudo, a literatura neurocientífica contém poucas derivações interteóricas explícitas, se é que possui alguma. Para Schaffner (1993), a redução deve ser apenas um ideal regulador, devendo os cientistas se contentarem com descrições fragmentárias e parciais. No entanto, como aponta Craver (2007), as reduções estão muito distantes da prática neurocientífica para servirem como ideal regulador. O que mais confere unidade à neurociência é a colaboração entre pesquisadores de diferentes campos para elaborarem explicações mecanísticas multinivelares. A explicação tende a ser correta se ela for consistente com técnicas e pesquisas teórica e causalmente independentes. Isso será discutido mais profundamente no decorrer do capítulo.

2.1.4 Critérios de adequação de uma abordagem explicativa em neurociência

Craver delimita três critérios de adequação de uma abordagem filosófica da explicação em neurociência. Em primeiro lugar, ela deve ser adequada do ponto de vista descritivo. Com isso, ele quer dizer que o modelo de explicação deve realmente refletir como a ciência é, e não a ideia do filósofo de como a ciência deveria ser.

Em segundo lugar, a abordagem deve ser capaz de distinguir explicação de outros tipos de conquista científica. Por exemplo, deve ser capaz de diferenciar uma explicação de uma simulação do fenômeno ou uma categorização de suas partes.

Por último, a abordagem explicativa deve oferecer critérios para avaliar explicações. Isso quer dizer que ela não deve apenas descrever como são as explicações em neurociência, mas fornecer normas para o que boas explicações devem ser.

2.2 Explicação e relevância causal

A visão proposta por Craver, que ele denomina mecanístico-causal, opõe-se a outras abordagens da explicação científica, incluindo o modelo nomológico-dedutivo de Carl Hempel e o modelo unificacionista de Philip Kitcher. Ela possui alguns aspectos em comum com a abordagem mecânico causal proposta por Salmon (1984), especialmente por considerar que explicações deve mostrar como os fenômenos se situam na estrutura causal do mundo. Explicações não são, portanto, argumentos, mas sim características objetivas do mundo.

Considere-se novamente o exemplo da liberação de neurotransmissor pelo neurônio. Nesse caso, o *explanandum* é a própria liberação de neurotransmissores na fenda sináptica. O *explanans* é (em parte) o mecanismo que gera o influxo de Ca^{2+} no axônio terminal. O trabalho que levou à descoberta do mecanismo explorou a relação entre esse íon e a liberação de neurotransmissores. Os cientistas descobriram que o aumento da concentração de Ca^{2+} intracelular (injetando-o diretamente no axônio terminal, por exemplo) está relacionado com um aumento da liberação de neurotransmissor. Da mesma forma, a diminuição da sua concentração intracelular (por exemplo, diminuindo a sua concentração extracelular e impedindo o seu influxo) diminui ou impede a liberação do neurotransmissor. Outros fatores que ocorrem junto com o potencial de ação, como o aumento da concentração intracelular de Na^{2+} , não são, por si só, relevantes para a explicação do fenômeno.

Esse exemplo contém algumas normas implícitas do que deve ser uma boa explicação em neurociência. A primeira delas é que, para determinar quando um fator possui relevância explicativa, é necessário que ele preceda o explanandum, embora isso não seja suficiente. Por exemplo, caso o aumento de Ca^{2+} intracelular ocorresse após a liberação do neurotransmissor, ele não poderia ser considerado relevante para a explicação do fenômeno. Essa é a segunda norma: causas e efeitos não podem ser confundidos. A causa explica o efeito, não o oposto.

A terceira norma é que um efeito não pode ser explicado por outro com uma causa comum. O aumento da concentração de Ca^{2+} intracelular tem vários outros efeitos além da liberação de

neurotransmissor, mas a maior parte deles é irrelevante para a explicação do fenômeno.

A quarta norma é que fenômenos causalmente irrelevantes não explicam o fenômeno. Por exemplo, o influxo de Na^{2+} e efluxo de K^{+} ocorrem simultaneamente ao potencial de ação e, portanto, previamente à liberação do neurotransmissor, mas não podem ser usados para explicar a liberação do neurotransmissor.

A quinta norma é um pouco menos evidente. A relevância estatística é um guia relativamente confiável, mas falho, para se avaliar a relevância explicativa. Em neurociência, há casos em que os fatores explicativos não aumentam a probabilidade do *explanandum*. Por exemplo, poderia haver um mecanismo alternativo de liberação do neurotransmissor que não fosse mediado por Ca^{2+} e que fosse desencadeado quando o primeiro falhasse. Nesse caso, o influxo de Ca^{2+} não aumenta a probabilidade de ocorrência do evento, uma vez que o mecanismo alternativo seria disparado caso o primeiro falhasse. Ainda assim, considera-se que ele é relevante para a explicação do fenômeno. Da mesma forma, os fatores explicativamente relevantes para um fenômeno não precisam torná-lo provável. De fato, a relação entre a ocorrência do potencial de ação, juntamente com a liberação de Ca^{2+} e a liberação de neurotransmissor é estocástica. Apenas dez a vinte por cento deles resultam em liberação do neurotransmissor.

Em suma, as cinco normas para explicações em neurociência são:

- (E1) Sequências temporais não são explicativas;
- (E2) Causas explicam os efeitos, e não vice versa;
- (E3) Efeitos independentes de causas comuns não explicam uma à outra;
- (E4) Fenômenos causalmente irrelevantes não são explicativos;
- (E5) Causas não precisam tornar os efeitos prováveis para explicá-los.

2.2.1. O modelo ND frente às normas da explicação em neurociência

Como vimos no primeiro capítulo, o modelo ND, afirma que explicações científicas são argumentos compostos por premissas que descrevem leis da natureza e condições antecedentes e uma conclusão que descreve o fenômeno a ser explicado. O poder explicativo, segundo essa visão, advém da capacidade de mostrar que o fenômeno descrito na conclusão seria esperado de acordo com as leis da natureza. O *explanandum* é a conclusão de um argumento sólido. Assim, explicação e previsão andam juntas; a única diferença é que, nas explicações, a conclusão do argumento é presumida verdadeira. Uma vez que já discutimos as limitações gerais desse modelo no capítulo, limitaremos-nos a mostrar como os seus problemas relacionam-se às normas E1 a E5.

Uma das falhas do modelo ND é sua incapacidade de distinguir entre leis da natureza que são genuinamente explicativas de generalizações acidentais, que não o são. Generalizações que descrevem meras sequências temporais (E1), correlações (E2) ou efeitos comuns à mesma causa (E3) não são explicativas. O modelo nomológico dedutivo pode cumprir o papel de um ideal regulador, mas isso depende de uma adequada abordagem de leis.

Outro problema do modelo D é sua incapacidade de distinguir condições antecedentes que são relevantes das que não o são (E4). Por exemplo, no caso do potencial de ação, o aumento da voltagem de membrana, mas não o aumento da concentração intracelular de Na^{2+} , é relevante para explicar a abertura dos canais de Ca^{2+} . Entretanto, a generalização que liga o aumento de Na^{2+} e a liberação de neurotransmissores existe, realmente, embora seja apenas uma correlação não causal. Nesse caso, o modelo nomológico-dedutivo por si só se revela incapaz de distinguir entre a condição antecedente relevante (o aumento da voltagem de membrana) e aquela que é irrelevante (o aumento da concentração intracelular de Na^{2+}). Isso ocorre porque, ao considerar que a explicação é um argumento, o modelo ND aceita que a adição de premissas irrelevantes não faça diferença para a explicação, uma vez que a força de um argumento não diminui com elas. Ao contrário, uma boa explicação neurocientífica necessariamente precisa excluir antecedentes irrelevantes, considerando somente os fatores causais como parte da explicação.

Embora para alguns filósofos, como Salmon, esses problemas sejam fatais para o modelo nomológico-dedutivo, existem filósofos céticos da causalidade que se posicionam diferente. Um deles é John Norton (2003) que, seguindo Bertrand Russel (1913), defende que as normas E1 a E4 são mera noções populares⁶ que precisam ser abandonadas à medida que a ciência amadurece. Embora em alguns “ambientes favoráveis”, como algumas áreas da biologia ou fisiologia, a regularidade de muitos fenômenos evoque a ilusão de uma relação causal, eles argumentam que, em última análise, tais relações causais são apenas generalizações sustentadas por leis que não são causais em um nível fundamental.

Craver argumenta que, mesmo que essa postura cética seja o caso, a biologia é um desses ambientes favoráveis em que alguns fenômenos dependem de outros e no qual nem toda correlação é igualmente explicativa. Ao contrário de físicos fundamentalistas, que podem explicar vários fenômenos exaustivamente através de equações diferenciais, neurocientistas podem refinar suas observações e compreender como as regularidades descritas em tais equações se dão. Não há qualquer motivo para tomar a trajetória da física como um padrão a ser projetado no desenvolvimento de todas as ciências.

Ademais, deve-se levar em conta que um dos objetivos da neurociência é o tratamento de

6 O termo original usado por Craver é “folk notion” (Craver, 2007, p. 37).

doenças, a melhora da função cerebral, a prevenção do declínio cognitivo e outras formas de manipulação, como já foi dito anteriormente. Por esse motivo, é essencial distinguir relações causais de meros correlatos, fatores irrelevantes e predecessores temporais.

Um último problema para o modelo ND para as explicações em neurociência é que ele se apóia na tese da expectabilidade nômica, ou seja, de que explicar um evento é mostrar como ele seria esperado com base nas leis da natureza e as condições antecedentes. Entretanto, em neurociência, há vários fenômenos que ocorrem de forma estocástica. Por exemplo, apenas dez a vinte por cento dos potenciais de ação levam à liberação de neurotransmissores; canais iônicos abrem-se frequentemente em situações improváveis; apenas um quinto dos casos de sífilis não tratada terminam em paralisia geral progressiva⁷. Explicar um fenômeno neurocientífico não é, portanto, igual a mostrar que ele deve ser esperado com base em condições antecedentes.

2.2.2 O modelo unificacionista e as explicações em neurociência

Um dos problemas da abordagem unificacionista para as explicações em neurociência relaciona-se à abordagem causal usada por Kitcher. Ele assume uma visão cética sobre a causalidade, defendendo que, rigorosamente, não existem relações causais no mundo. Ainda assim, ele aceita que os pontos E1 a E5 são importantes restrições de abordagens causais. Para ele, a sistematização mais unificada de K deve incluir padrões de argumentos que satisfaçam tais restrições. Ele se utiliza de uma complexa argumentação, que nem sempre tem sucesso, para que o modelo unificacionista não admita que efeitos expliquem suas causas e para lidar com efeitos pouco prováveis, como a abertura de canais iônicos e a liberação de neurotransmissor pelo neurônio.

Além disso, o modelo unificacionista não permite a distinção entre explicações neurocientíficas boas e ruins. Para avaliar explicações, de acordo com esse modelo, é necessário avaliar o próprio conjunto de argumentos que forma E(K). Contudo, uma vez que esse modelo depende de verdades contingentes e que elas mudam com o tempo, o que hoje parece um progresso em direção a E(K) pode parecer supérfluo sob a perspectiva da ciência futura.

2.2.3 O modelo de Hodgkin e Huxley

Muitas das explicações mais relevantes em neurociência parecem exemplificar o modelo nomológico dedutivo ou o modelo unificacionista. O modelo de Hodgkin e Huxley (modelo HH –

⁷ A paralisia geral progressiva é uma manifestação tardia da sífilis, causada por danos no sistema nervoso central. Ela inclui sintomas psiquiátricos e neurológicos e, se não tratada, leva ao óbito em poucos anos.

anexo 1) é um exemplo.

Weber (2005), por exemplo, defende que esse é um caso paradigmático de fenômeno biológico sendo explicado por leis físico-químicas. Realmente, as equações do modelo HH permitem prever como a corrente varia sob diversas intervenções experimentais, podem ser usadas para simular atividades eletrofisiológicas de células nervosas, permitem inferir o valor de variáveis não medidas e constituem fortes evidências de que um mecanismo envolvendo correntes iônicas poderia dar conta da forma do potencial de ação.

Entretanto, isso não quer dizer que elas *expliquem* o potencial de ação. As equações nada dizem sobre os mecanismos que promovem a mudança da condutância da membrana, que permitem o fluxo de íons e que coordenam as mudanças de condutância que dão ao potencial de ação a sua forma peculiar. Elas não mostram *como* a membrana muda sua permeabilidade. Para explicar, esse modelo precisaria ser complementado por uma descrição objetiva dos mecanismos que envolvem o potencial de ação, ou seja, de como as entidades e atividades estão organizadas para produzir o potencial de ação. Para Craver, ele se parece muito mais com o modelo ptolomaico do movimento dos planetas do que com o modelo gravitacional de Newton.

Os próprios autores (Hodgkin e Huxley, 1952) afirmam de forma explícita que o seu modelo não é explicativo. Para eles, o modelo consiste apenas em uma descrição empírica das mudanças da permeabilidade do sódio e do potássio no tempo. Ele insistem que equações muito diferentes poderiam igualmente prever o comportamento elétrico da membrana e reconhecem que, embora as equações sejam passíveis de interpretação física, o seu sucesso não é evidência do mecanismo que eles tinham em mente quando a formularam. Elas consistem apenas em uma descrição temporal das mudanças de permeabilidade. Como argumenta Bogen (2005), as equações do modelo HH são apenas descrições matemáticas de um esboço de mecanismo, que resumem os resultados de experimentos conduzidos em décadas.

De fato, a elucidação do mecanismo do potencial de ação só ocorreu muitos anos depois, com os trabalhos de Hille (1984) e Hille et al (1999). Inicialmente, Hodgkin e Huxley propuseram que o fluxo de canais iônicos ocorreria através de um mecanismo de translocação, mas admitiram não possuir quaisquer evidências desse mecanismo. Somente na década de 60 começou a surgir a noção de um canal iônico, mas apenas com o isolamento das proteínas que formam esses componentes e a descrição de como eles são ativados e inativados, de como ocorre o fluxo iônico através deles durante o potencial de ação, ou seja, de como se dão as relações causais entre eles é que foi possível completar o que antes era apenas um esboço do mecanismo. Por esse motivo, o modelo de Hodgkin e Huxley não pode ser considerado um triunfo dos modelos de explicação nomológico-dedutivo ou unificacionista.

2.3 Relevância causal e manipulação

No tópico anterior, discutimos as cinco normas (E1 a E5) que uma boa explicação em neurociência deve satisfazer. Essas normas pressupõem uma visão da estrutura causal do mundo que seja compatível com elas. Ainda que a abordagem mecanístico-causal possua algum comprometimento metafísico ao propor uma concepção ôntica da explicação científica, onde o cientista deva *descobrir* como o mecanismo se situa numa estrutura causal do mundo, ele não se propõe a uma análise metafísica da causação em si, que a descreva em termos de conceitos não causais. Não existe uma busca pela resposta à questão humeana da conexão entre a causa e o efeito, como fizeram Salmon (1984, 1998), Dowe (2000) e Glennan (1996). Afinal, muitas causas em neurociência não estão conectadas com os seus efeitos, como nos casos de omissão e prevenção, que examinaremos adiante. Para essa abordagem, basta que a causação seja compreendida como as várias *atividades* que os cientistas descrevem nas suas teorias, tais como colisão, difusão, atração, repulsão, magnetismo, etc.

Vários fenômenos em neurociência, como a sensitização de longo prazo (anexo 2), possuem algumas características particulares. Em primeiro lugar, eles *possuem uma abrangência limitada*. Nem todas as células, sinapses, nem mesmo todas as sinapses glutamatérgicas exibem a sensitização de longo prazo, que varia de acordo com vários fatores, como organismo, região cerebral e estágios do desenvolvimento. Segundo, eles ocorrem de maneira *estocástica*. No nosso exemplo, mesmo nas sinapses em que ocorre, não ocorre sempre. Isso se deve, pelo menos em parte, à sua *fragilidade mecânica*, ou seja, à sua suscetibilidade a diversas intervenções externas. Por fim, generalizações sobre o fenômeno são *historicamente contingentes*, não são verdades atemporais sobre o cérebro e seus componentes, mas explicações parciais sobre produtos de mecanismos agrupados juntos ao longo da evolução por seleção natural e aperfeiçoados com o desenvolvimento.

Apesar disso, as generalizações que descrevem a sensitização de longo prazo são relações causais em mecanismos que realmente funcionam: não são meras sequências temporais; relacionam causas a efeitos, não vice versa; não descrevem relações entre efeitos de causas comuns; e descrevem relações entre eventos relevantes, apenas. É necessário, portanto, uma abordagem de causalidade de permita lidar com as características peculiares dos fenômenos neurocientíficos. A seguir, examinaremos como isso limita o uso de abordagens causais clássicas, propondo uma alternativa que permita lidar com essas características.

2.3.1 Causação como transmissão

As abordagens de causação feitas por Salmon (1984, 1997) e por Dowe (1992, 2000) contêm uma visão amplamente aceita, embora muitas vezes implícita, de que a causação envolve o contato entre objetos e troca ou transmissão de algo entre eles. Salmon desenvolveu duas abordagens da causação: a da transmissão de marca (TM) e a conservação de quantidade (CQ). A segunda foi formulada com a tentativa de resolver alguns problemas da primeira e acabou substituindo-a.

As duas abordagens são embasadas em dois conceitos centrais: os *processos causais* e as *interações causais*. Processo, aqui, tem o sentido de coisas que exibem consistência de características através do tempo. Isso difere de uma noção de processo como eventos ou ocorrências estendidas, como processos computacionais e processos de produção. Por exemplo, uma molécula cruzando uma sinapse, um íon Ca^{2+} entrando na célula e uma sombra se movendo no chão à medida que um carro se move na estrada são todos processos. De acordo com a TM, os processos causais são aqueles capazes de transmitir uma marca, ou seja, uma mudança de alguma característica de um processo quando ele se intercepta com outro. É isso o que distingue processos de pseudo-processos, que não transmitem marca.

Tanto na TM como na CQ, explicar um fenômeno é situá-lo em umnexo causal. No entanto, isso não é suficiente para cumprir as exigências de uma abordagem causal em neurociência. Em primeiro lugar, a TM se mostra incapaz de lidar com interações transitórias envolvendo marcas que ocorrem apenas durante as interações entre os processos. Por exemplo, quando o glutamato se liga ao NMDA, esse receptor muda de conformação e expõe o canal através da membrana, retornando à sua conformação original tão logo o glutamato se desligue. Não existe, nesse caso, nem transmissão de marca, nem de quantidades causais.

O segundo problema foi percebido pelo próprio Salmon, que o considerava grave o suficiente para comprometer a TM. Quando a desenvolveu, ele pretendia tratar da causação como um fenômeno empírico, aberto à investigação com métodos da ciência física. Isso envolveria, para ele, uma abordagem causal livre de contrafactuais. Contudo, ele próprio se deu conta, posteriormente, de que a TM apela tacitamente a contrafactuais. Foi esse problema que o levou a formular a abordagem CQ.

A abordagem CQ analisa a causação de uma forma que se encaixa bem em explicações de ciências fundamentais, uma vez que a troca de quantidades conservadas são feitas nesse nível. Embora ela consiga resolver o problema dos contrafactuais, ela perde a abrangência da MT que possibilitava o seu uso para a descrição de interações causais em neurociência. Embora explicações

em neurociência não violem as leis da conservação, elas não fazem apelo direto a trocas de quantidades conservadas, uma vez que seus processos ocorrem em um nível não fundamental.

Especificamente, a relevância causal, (critério E4), não pode ser analisada em termos de trocas de quantidades conservadas apenas. O nexos causal é uma rede de processos e interações causais, da qual apenas os relevantes precisam ser selecionados. É necessário selecionar quais os fatores fazem diferença para o efeito, e não apenas quais as marcas são transmitidas ou quais as trocas de quantidades são feitas. Tanto a TM como a CQ são incapazes de distinguir entre marcas que fazem e que não fazem diferença. Esse problema só poderia ser resolvido com o uso de contrafactuais.

Outra dificuldade da TM e da CQ é que elas não dão conta de explicações que apelam para fatores causais negativos, essenciais em explicações neurocientíficas. Ausências não portam ou trocam quantidades conservadas. Não são processos, coisas e não exibem consistência de características através do tempo. Mesmo assim, existem vários fenômenos em neurociência que são unanimemente aceitos por neurocientistas como causados por ausências. Por exemplo, a ausência do bloqueio de Mg^{2+} causa a entrada de Ca^{2+} na célula ou, o reverso, a localização do Mg^{2+} no canal impede que as concentrações de Ca^{2+} elevem-se mesmo que o glutamato se ligue ao receptor NMDA. Como argumenta Schaffer (2004), uma visão de causação que não inclui causas negativas está severamente em conflito com o discurso de senso comum sobre causas, com julgamentos científicos sobre o que causa o quê e com aplicações teóricas do conceito de causa.

Causas negativas, contudo, não são aceitas consensualmente entre os filósofos. Uma das razões é metafísica: ausências, como nada, não têm poder causal. Craver exime-se da discussão metafísica e orienta uma visão segundo a qual relações causais não são eventos, estados das coisas, propriedades ou valores de variáveis, por exemplo. Para ele, uma relação causal possui duas variáveis, causa e efeito, sendo a relação causal avaliada em termos do contraste entre elas. No caso da variável causal, o contraste ocorre entre o valor da variável em uma intervenção ideal e o valor que ela tem em condições de controle. No caso da variável de efeito, o contraste é entre o valor da variável em condições controle e seu valor em condições experimentais.

Outro problema das causas negativas é que, pelo menos em primeira análise, elas seriam muitas e, na maior parte dos casos, irrelevantes para a explicação, inclusive em neurociência. Na tentativa de resolver esse problema, Dowe (2000, 2004) propõe chamar de causação* as causas que não envolvem conexão física, como as causas negativas, limitando o termo “causação” às causas que envolvem conexões físicas. Craver acredita que essa estratégia apenas muda o problema de lugar. Ao invés disso, ele oferece uma solução que ele próprio chama de pragmática, propondo que alguns casos de causação sejam excluídos de textos explicativos por serem improváveis demais ou

anormais. Craver sustenta que causas negativas possuem um papel importante demais nas explicações em neurociência para serem ignoradas. Segundo ele, a distinção entre causas positivas e causas negativas é meramente psicológica.

2.3.2 Causação e conexão mecânica

Craver admite que a sua noção de causação tem forte influência da de Glennan (1996). Contudo, os autores possuem objetivos diferentes. Enquanto Glennan pretendia responder aos questionamentos céticos de Hume quanto ao poder oculto que liga a causa ao efeito, Craver, como já vimos, pretende desenvolver uma abordagem causal adequada à noção de mecanismos em neurociência.

Glennan afirma que o poder causal oculto que liga uma causa X a um efeito Y é um mecanismo que liga X a Y. Aqui, mecanismo pode ser compreendido como um sistema complexo que produz o seu comportamento através da interação entre várias partes de acordo com leis causais diretas. O mecanismo, nesse caso, seria a conexão secreta que liga eventos causalmente ligados em um nível não fundamental. Desse modo, o problema humeano em um nível não-fundamental pode ser resolvido. Ainda assim, ele admite que, em níveis fundamentais, a questão humeana persiste, uma vez que o que se pode observar são apenas regularidades, sendo impossível oferecer uma explicação de como tais regularidades se dão.

Para Craver, o principal problema da abordagem de Glennan é a contradição entre sua posição declaradamente anti-fundamentalista e a sua tentativa de analisar a causação em termos de mecanismos de nível mais baixo. Ao mesmo tempo em que ele rejeita a noção de uma cola metafísica que liga os correlatos causais em um nível não fundamental, ele aceita que, em níveis fundamentais, isso ocorra. Craver acredita que, ao fazer isso, Glennan simplesmente empurra a ignorância sobre a natureza das causas um pouco mais abaixo.

A única maneira de encerrar essa questão é oferecendo uma resposta para a questão da conexão causal em níveis fundamentais. Glennan propõe que a regressão termine em leis fundamentais existentes em níveis mais baixos, mas não desenvolve uma abordagem adequada que permita distinguir leis de acidentes. Isso cria um paradoxo com o restante da sua teoria, uma vez que uma conexão causal necessária entre X e Y (em um nível não-fundamental) não poderia ser composta por relações na qual não existe qualquer necessidade (em um nível fundamental).

Craver acredita que, na sua tentativa de responder à questão humeana, Glennan se desviou de abordar mais profundamente uma hierarquia de mecanismos. Por isso, ele propõe uma abordagem capaz de satisfazer E1 a E5 sem enveredar pelo que ele chama de “obscuro

fundamentalismo”. Isso será explorado no próximo item.

2.3.3 Causação e manipulação

Para Craver, relevância causal, explicação e controle andam juntos, particularmente nas ciências biomédicas, cuja motivação não é apenas curiosidade intelectual. A busca por boas explicações, nesse caso, vem impregnada do desejo de compreender onde e como intervir no mundo na tentativa de mudá-lo. Os próprios testes usados por neurocientistas refletem essas preocupações, uma vez que eles não se limitam a revelar as partes de um mecanismo, mas também mostrar como mudar o seu comportamento de forma previsível. Para a neurociência, mostrar que um estágio do mecanismo produz outro, ou ao menos é relevante para ele é, em parte, o mesmo que dizer que é possível manipular o segundo através de intervenções no primeiro.

A abordagem causal usada por Craver é muito semelhante à de Woodward (2003). Esse autor defende que relações causais se distinguem das demais porque são potencialmente exploráveis para manipulação e controle. Especificamente, uma variável X é relevante para uma variável Y , nas condições W , se uma intervenção ideal I em X nas condições W mudam o valor de Y ou a probabilidade de distribuição do seu valor.

Aqui, intervenção de X com relação a Y tem o sentido de uma manipulação que muda o valor da variável. A intervenção não precisa ser humanamente possível, apenas potencialmente possível. Pode ser que não se saiba como intervir, ou que os instrumentos existentes seja insuficientes, ou que X seja muito grande ou muito pequeno, desde que a relação entre X e Y seja passível de manipulação. Diz-se que uma intervenção de X com relação a Y é ideal quando o valor de Y muda somente de acordo com o valor de X . Formalmente, isso pode ser representado da seguinte maneira:

(I1) I não muda o valor de Y diretamente;

(I2) I não muda o valor de um intermediário causal entre X e Y , exceto pela mudança do valor de X ;

(I3) I não está correlacionado com uma outra variável M que é a causa de Y ;

(I4) I age como uma chave que muda o valor de X a despeito das outras causas de X , U .

As generalizações explicativas que descrevem essa relações causais são estáveis (ou invariantes, nos termos de Woodward), mas não necessariamente universais. A relação entre a variável de causa e a variável de efeito ocorre sob uma gama de condições, que incluem condições

de estímulo, condições intermediárias e condições de fundo. As generalizações explicativas em neurociência podem variar consideravelmente, desde que permaneçam estáveis nas condições relevantes para um certo interesse explicativo. Essa flexibilidade é adequada para a neurociência, cujas generalizações causais são limitadas em abrangência, mecanisticamente frágeis e historicamente contingentes.

Evidentemente, ser capaz de manipular um fenômeno não é suficiente para explicá-lo. Porém, quanto maior a quantidade de possíveis manipulações de um fenômeno, mais completa é sua explicação. Textos explicativos mais profundos respondem a mais perguntas sobre o comportamento do sistema em uma série de intervenções do que textos mais rasos.

A abordagem manipulacionista satisfaz os critérios E1 a E5. Ela apresenta vantagens em relação a outras abordagens causais, pois é capaz de diferenciar casos de meras sequências temporais como relações de causa e efeito (E1). Como não é possível mudar eventos do passado através da intervenção em eventos presentes, ela também satisfaz E2. Em E3, a abordagem manipulacionista também leva vantagem. Se a relação entre duas variáveis é de mera correlação, então não é possível manipular uma variável intervindo no valor da outra. Essa abordagem também é capaz de distinguir entre propriedades e interações relevantes e irrelevantes (E4), uma vez que a intervenção naquelas irrelevantes não é capaz de produzir mudanças na variável de efeito. Por fim, uma vez que a abordagem manipulacionista requer apenas que a intervenção seja capaz de mudar a probabilidade de distribuição do evento, sem demandar que ele se torne provável, ela também é capaz de dar conta de E5.

Além de ser capaz de lidar com todas os critérios acima, a abordagem manipulacionista também se revela capaz de acomodar a causação por omissão e prevenção. Dadas as variáveis X e Y, considere-se que elas podem assumir os valores “ausente” e “presente”. Se ocorre mudança em Y a partir da mudança de X de “presente” para “ausente”, pode-se dizer que a ausência de X causou uma mudança em Y. Analogamente, se a mudança de X é capaz de mudar o valor de Y de “presente” para “ausente”, é possível afirmar que a mudança em X preveniu Y.

A habilidade da abordagem manipulacionista para satisfazer E1 a E5 torna-a a mais adequada para as explicações em neurociência. Quando as relações de manipulabilidade são conhecidas, é possível identificar quais as intervenções fazem diferença para o *explanandum*, permitindo prever e manipular o seu comportamento sob uma variedade de condições.

2.4 As normas da explicação mecanística

Craver afirma que as explicações mecanísticas podem ser de vários tipos. Explicações

etiológicas explicam um evento ao descrever suas causas antecedentes. Por exemplo, a desidratação é uma parte da explicação etiológica da sede. Explicações mecanísticas podem ser também constitutivas ou componenciais, quando descrevem o mecanismo subjacente de um fenômeno. O receptor NMDA, por exemplo, é parte da explicação constitutiva do PLD. Existem duas tradições de pensamento das explicações constitutivas: a tradição redutiva e a tradição de sistemas.

Segundo a tradição redutiva, explicações constitutivas são uma espécie de modelo nomológico-dedutivo entre teorias de níveis diferentes. A explicação é feita construindo declarações de identidade entre termos-tipo de nível alto e termos tipo de nível baixo e, a partir disso, derivando as leis das teorias de nível mais alto a partir das leis de nível mais baixo. Ela possui um comprometimento epistêmico de que a compreensão é uma expectativa racional do *explanandum*, com base no *explanans*. Além disso, ela oferece um ideal regulador para a explicação. De acordo com essa tradição, seria possível derivar o *explanandum* do *explanans* quando a explicação estivesse completa.

Atualmente, segundo Craver, existem poucos defensores dessa corrente. O principal motivo é que tipos de nível mais alto são multiplamente realizados pelos tipos de nível baixo, de tal forma que não existe identidade entre os termos-tipo de nível mais alto e de nível mais baixo. Ademais, as explicações que realmente existem em neurociência não se parecem em nada com o propõe a tradição redutiva. Além disso, esse modelo possui as mesmas limitações do modelo nomológico-dedutivo, discutido no item 2.2.2.

A tradição de sistemas formula explicações decompondo sistemas em suas partes e mostrando como essas partes se organizam de tal forma que exibem o *explanandum*. Como diz Dretske (1994), citado por Craver, “se você não pode construí-lo, então você não sabe como ele funciona”. Nesse sentido, a tradição de sistemas segue uma espécie de ideal de engenheiro do que realmente é necessário para entender como algo funciona.

Na visão de Craver, uma das formulações mais rigorosas da tradição de sistemas é a análise funcional de Cummins. Para ele, o *explanandum* é uma capacidade ψ de um sistema S . A ψ -ação de S é explicada analisando-a em subcapacidades $\{\phi_1, \phi_2, \dots, \phi_n\}$, mostrando que ψ é produzido através da execução programada das subcapacidades.

Assim, explicações de sistemas consistem em mostrar como algo funciona, e não como seu comportamento pode ser derivado de leis mais fundamentais. Uma de suas vantagens é que não há qualquer problema com a múltipla realização do *explanandum*. Como ela rejeita a noção de que explicações são argumentos, não requer que termos-tipo de nível mais alto sejam traduzidos em termos-tipo de nível mais baixo. Explicações de sistemas estão muito mais próximas explicações que neurocientistas buscam e constroem.

2.4.1 Uma alternativa mecanística

Craver defende que o seu modelo de explicação é mecanístico e normativamente adequado. Ele almeja formular um modelo normativo que permita distinguir explicações do tipo “como-possivelmente” de explicações “como-verdadeiramente”, além de distinguir esboços de mecanismos de esquemas (*schemata*) de mecanismos. Esse modelo tem o intuito de ser um ideal regulador, segundo o qual explicações constitutivas devem descrever todas e somente as entidades, atividades, propriedades e características organizacionais que são relevantes para o fenômeno a ser explicado. Para elaborar seu modelo, ele analisa separadamente alguns aspectos da explicação mecanística. São eles: (i) a natureza do *explanandum*; (ii) a relação constitutiva entre o fenômeno e seus componentes; (iii) a diferença entre componentes reais e ficções úteis; (iv) a natureza das capacidades e atividades; (v) a natureza da organização mecanística; (vi) a natureza da relevância explicativa constitutiva.

Craver identifica duas distinções normativas implícitas nas práticas de construção e avaliação das explicações mecanísticas. A primeira delas é distinção entre modelos *como-possivelmente* e modelos *como-verdadeiramente*. Modelos *como-possivelmente* são hipóteses frouxamente concisas sobre o tipo de mecanismo adequado para produzir o *explanandum*. Eles mostram como um conjunto de partes e atividades pode se organizar para produzir um fenômeno, sem se preocupar com a real existência das partes conjecturadas nem sua capacidade de engajar-se nas atividades atribuídas a elas. Tais modelos são heurísticamente úteis na construção e exploração de mecanismos possíveis, mas não são explicações adequadas. Por outro lado, modelos *como-verdadeiramente* descrevem os reais componentes, atividades e características organizacionais do mecanismo que produz o fenômeno. Eles mostram como o mecanismo funciona, e não simplesmente como ele poderia funcionar. Entre os dois extremos encontram-se modelos *como-plausivelmente*, que mais ou menos consistentes com as restrições aos seus componentes, atividades ou organizações.

A segunda distinção proposta por Craver é entre *esboços de mecanismos* e *modelos completos de mecanismos*. Um *esboço de mecanismo* é um modelo incompleto do mecanismo. Ele descreve algumas partes, atividades ou características organizacionais, mas deixa falhas. Eventualmente, essas falhas são mascaradas por *termos de preenchimento*, como (ativação, inibição, codificação, causa, produção, etc) que dão a ilusão de que a explicação está completa, podendo constituir obstáculos ao progresso futuro. Entre os esboços de mecanismo e as descrições completas há um continuum de *esquemas de mecanismos*, que são apenas parcialmente compreendidos. Para

ilustrar essas distinções, Craver usa o exemplo da descoberta do potencial de ação, que examinaremos a seguir.

2.4.2 A descoberta do potencial de ação

A história da descoberta do potencial de ação fornece um exemplo de explicação mecanística bem sucedida, ilustrando as distinções discutidas acima e as normas implícitas na prática de construção de explicações mecanísticas constitutivas. Hodgkin e Huxley construíram um modelo matemático do potencial de ação, como já foi discutido anteriormente. Essa primeira tentativa de explicação, contudo, tratava-se de um esboço de um mecanismo. Hodgkin e Huxley conheciam algumas entidades e atividades. Sabiam, por exemplo, que o potencial de ação é produzido pela despolarização da célula, que ele envolve o movimento de íons através da membrana e que o seu formato pode ser explicado pela ativação e inativação da condutância de membrana para determinados íons. Contudo, “ativação” e “inativação” eram apenas termos de preenchimento.

Inicialmente, os autores propuseram que algumas partículas de “ativação” poderiam enfraquecer a integridade da membrana, ainda que admitindo não possuir qualquer evidência que favorecesse essa hipótese em relação a outra semelhante. Eles usaram um diagrama formado por um capacitor e três circuitos em paralelo para representar as correntes que compunham a equação total da corrente. Cada um dos circuitos continha, em série, uma bateria, que representa o potencial de equilíbrio de cada íon e um resistor, que é o inverso da condutância do íon. O modelo mostrava que as mudanças coordenadas às correntes de Na^+ e K^+ poderiam explicar as características do potencial de ação.

Embora esse modelo oferecesse pistas valiosas sobre o mecanismo do potencial de ação, ele era apenas um esboço, principalmente devido ao uso dos termos de preenchimento “ativação” e “inativação”. Ele encorajou o avanço de pesquisas sobre a biofísica da membrana e a busca por canais iônicos, mas já nos anos 1970 ainda se falava em canais iônicos como uma metáfora, como termos de preenchimento.

Somente com os trabalhos de Armstrong (1981) e Hille (1992), entre outros, foi construído um modelo, hoje aceito, segundo o qual as mudanças de condutância nos potenciais de ação podem ser explicadas pela abertura e fechamento coordenados de canais transmembrânicos de Na^+ e K^+ . Isso elevou os canais íon-específicos acima do status de termos de preenchimento. Entretanto, embora esse modelo acrescentasse alguns detalhes ao modelo de Hodgkin e Huxley, ele ainda permanecia um esboço, principalmente devido ao uso dos termos de preenchimento “ativação” e

“inativação”. Permanecia sem explicação a questão de *como* os canais faziam. Por isso, Hille propôs, então, um conjunto de modelos como-possivelmente dos canais, que possuíam partes e atividades diferentes, organizadas de formas diferentes.

À medida que surgiram os resultados dos experimentos envolvendo o sequenciamento de proteínas dos canais, a reconstrução da sua estrutura tridimensional, a identificação das regiões hidrófilas e hidrófobas e o comportamento dos canais em diversos tipos de intervenções elétricas, farmacológicas e genéticas, mais conhecimentos sobre a estrutura e o funcionamento dos canais de Na⁺ foram agregados. Hoje, sabe-se, por exemplo, que o canal de Na⁺ é composto por quatro subunidades com seis regiões transmembrânicas cada, num arranjo helicoidal chamada α -hélice. Sabe-se também que a abertura do canal é mediada pelo enfraquecimento do potencial positivo extracelular, causando a rotação da hélice e desestabilizando o balanço das forças que mantêm o canal fechado. Da mesma forma, o mecanismo de fechamento do canal pelo modelo conhecido como bola-e-corrente também é conhecido.

Embora ainda haja detalhes a serem desvendados, esse exemplo ilustra três aspectos das explicações mecanísticas: primeiro, que elas são moldadas pelo fenômeno *explanandum*, nesse caso, o potencial de ação; segundo, que a explicação é constitutiva, ou seja, o potencial de ação é explicado por referência às entidades (íons, canais, cadeias de proteínas) e atividades (difusão, mudanças de conformação); terceiro, que essas entidades são organizadas de forma a produzir o fenômeno. Em suma, modelos mecanísticos descrevem como entidades e atividades constitutivas são organizadas juntas em um fenômeno.

2.4.3 A caracterização do fenômeno *explanandum*

Craver defende que a principal exigência de uma explicação mecanística é que ela seja capaz de dar conta completamente do fenômeno *explanandum*. As fronteiras do mecanismo são delimitadas de acordo com o fenômeno que ele explica. A descrição de um fenômeno pode falhar de várias formas. Uma delas é quando o fenômeno é *ficício*, como foram as tentativas de explicar os mecanismo de vibração dos nervos conduzidas por David Hartley.

Podem ocorrer também erros taxonômicos, ou seja, de delimitação do fenômeno. Uma das formas que isso pode ocorrer é o que Craver chama de *erro de aglomeração* (*lumping error*), que consiste em considerar vários fenômenos distintos como um só. Durante muitos anos, por exemplo, as pesquisas sobre memória foram conduzidas como se ela fosse um fenômeno único. Mais tarde, tornou-se claro que a memória é composta de uma grande variedade de processos distintos e dissociados, o que tornou possível um avanço muito maior na área.

Outro tipo de erro taxonômico é o *erro de cisão*, que ocorre quando se assume que um fenômeno único são vários, como a tentativa de Franz Joseph Gall de dividir a mente em várias tendências, como o instinto homicida, ternura para com a prole, habilidade mecânica, facilidade com as cores, etc. É possível, inclusive, que todo um sistema taxonômico esteja equivocado.

Também podem ocorrer *erros por subespecificação* de um fenômeno. Fenômenos são multifacetados, que podem ser descritos com base em múltiplos aspectos. O potencial de ação, por exemplo, pode ser caracterizado em termos das suas condições precipitantes, da sua manifestação, (incluindo a taxa de elevação, o pico, a magnitude, a taxa de declínio, o período refratário), da suas condições inibitórias e moduladoras e dos seus subprodutos.

2.4.4 Componentes

Como foi dito anteriormente, Craver defende que explicações mecanísticas são do tipo constitutiva, ou componencial, explicando o comportamento do todo através do comportamento das suas partes. Embora seu ponto de vista guarde muitas semelhanças com a análise funcional de Cummins, existe um ponto crucial que distingue as duas.

Cummins defende uma forma de explicação analítica não-constitutiva. Para ele, tanto as capacidades como as subcapacidades (ψ e φ , respectivamente) são capacidades do sistema como um todo. No seu exemplo, para explicar a capacidade S de como um cozinheiro faz um bolo (ψ), apela-se às capacidade analisadas (φ) que, para ele, também são capacidades no nível do cozinheiro, como ler receitas, mexer a massa e acertar o ponto do sal. Craver concorda com Cummins que esses dois tipos de explicação devem ser distinguidos, mas toma um caminho oposto. Craver insiste que a ψ -ação de S não pode ser explicada meramente pelas subcapacidades de ψ , mas pelas capacidades $\{\varphi_1, \varphi_2... \varphi_n\}$ dos componentes de S $\{X_1, X_2... X_n\}$.

Essa diferenciação é crucial porque a distinção entre explicações como-verdadeiramente e como-possivelmente ocorre exatamente porque as últimas baseiam-se em partes componentes caracterizadas de forma incompleta. Alguns tipos de explicações funcionais descrevem um programa que poderia possivelmente produzir o fenômeno, dentre muitos outros que poderiam fazer o mesmo. Ou seja, em abordagens funcionais interpretativas, qualquer explicação como-possivelmente é suficiente, desde que as subcapacidades φ possam explicar a atividade ψ do sistema. Entretanto, isso não é suficiente para as explicações mecanísticas, inerentemente componenciais, nas quais os componentes reais devem, necessariamente, ser diferenciados de componentes fictícios.

Não existe um limite claro entre componentes reais e fictícios. Ainda assim, Craver descreve

algumas condições que componentes reais devem satisfazer, ajudando a distinguir entre explicações como-possivelmente e como-verdadeiramente. Em primeiro lugar, partes reais possuem um *conjunto estável de propriedades*. Os inicialmente especulativos canais iônicos, por exemplo, adquiriram o status de componentes quando foi possível identificá-los como proteínas, determinar a ordem dos seus aminoácidos, descobrir sua estrutura tridimensional, suas interações com neurotransmissores, agonistas e antagonistas, etc.

Segundo, as partes devem ser *robustas*. Suas características devem ser detectáveis através de instrumentos causalmente e teoricamente independentes. Canais iônicos podem ser isolados da membrana, purificados e sequenciados. Seu comportamento pode ser detectado através de técnicas de gravação e eles podem ser manipulados farmacologicamente, cristalizados e revelados por raio-X, bem como ser vistos através de microscópios eletrônicos. O uso de múltiplas técnicas de estudo de um componente diminui a chance de que as informações sejam enviesadas ou até mesmo equivocadas.

Em terceiro lugar, deve ser possível usar uma parte para *intervir* em outras partes e atividades, ou seja, manipular a entidade de maneira que produza mudança em outras entidades, propriedades e atividades. Por exemplo, é possível alterar o potencial de membrana, mudar a condutância ao Na^+ e abrir canais de K^+ através da manipulação de canais de Na^+ . Esse é um critério fundamental para estabelecer relações causais e explicativas entre componentes de um mecanismo e para distinguir entre componentes reais e fictícios.

Por último, o componente deve ser *fisiologicamente plausível*. Sua existência não deve ser possível apenas sob condições laboratoriais controladas, nem sob condições patológicas. Mais especificamente, a existência do componente deve ser demonstrada sob as condições relevantes para a explicação do fenômeno.

2.4.5 Atividades

Um dos méritos da análise funcional de Cummins e a tradição de sistemas, para Craver, é concentrar a atenção no tipo de explicação encontrado na psicologia e na neurociência. Contudo, Cummins não aborda questões filosóficas que, para Craver, são essenciais, como a distinção entre leis causais e correlações não causais, a exclusão de eventos não causais e o acesso independente a causas e efeitos. Para Craver, uma vez que análises funcionais são feitas de capacidades e que mecanismos são em parte constituídos por atividades e interações, uma abordagem constitutiva de explicação deve abordar essas questões.

Cummins descreve capacidades de um sistema como relações input-output que relacionam

condições antecedentes e manifestações subsequentes, as quais são regidas por “leis *in situ*”, ou seja, leis que regem um tipo de sistema com uma organização e constituição peculiares. Elas são equivalentes ao que Craver chama de generalizações mecanisticamente frágeis.

No entanto, existem alguns problemas que Cummins não deixa claros. Por exemplo, ele rejeita o modelo nomológico-dedutivo e afirma que leis não são explicativas, mas não demonstra de que maneira, então, suas leis *in situ* podem sê-lo, uma vez que não distingue as leis gerais do modelo nomológico dedutivo das leis especiais que regem as capacidades subjacentes na análise funcional.

Além disso, a análise funcional de Cummins não permite distinguir capacidades genuínas de pseudo-capacidades. O seu modelo de input-output pode ser entendido como mera sequência temporal, como correlações entre efeitos de causas comuns e como pares pseudocausa-efeito. Mesmo com o uso de contrafactuais, como proposto por Weber (2005), essa questão não é resolvida, uma vez que muitos contrafactuais não são explicativos. Dizer, por exemplo, que se o galo estava cantando então o dia estava amanhecendo não contribui para a explicação do fenômeno.

2.4.6 Organização

Textos explicativos precisam conter, além das partes e capacidades reais e completas, uma descrição de como esses componentes podem ser organizados. Embora Cummins defina organização como algo que possa ser descrito em um quadro ou programa, Craver argumenta que essa descrição é muito vaga, pois pode se referir a várias coisas além da organização.

Ele propõe que a organização, no sentido usado pelas explicações mecanísticas, pode ser compreendida contrastando mecanismos com meros agregados. Nesses últimos, o todo é literalmente a soma das suas partes. Isso quer dizer que as propriedades ou atividades do sistema podem ser explicadas pelas propriedades ou atividades das suas partes, de tal forma que não haja interações cooperativas ou inibitórias entre as partes e que o rearranjo, intersubstituição, adição, subtração, desagregação ou reagregação dessas partes não produza qualquer efeito sobre a atividade do todo.

Já nos mecanismos, os componentes estão organizados de três formas: ativa, espacial e temporalmente. Por esse motivo, as propriedades do sistema são mais que meras somatórias de propriedades, mas são o resultado das atividades e interações entre as entidades do mecanismo. Por esse motivo, as partes não podem ser organizadas, substituídas, removidas, acrescentadas, separadas e reagregadas sem que isso produza uma perturbação da atividade do sistema.

Explicações mecanísticas também se distinguem de explicações morfológicas, como é o

caso da explicação da forma de um cristal a partir das formas de suas moléculas e seu arranjo espacial. Enquanto nas explicações morfológicas a organização é meramente espacial e estática, nos mecanismos há uma organização ativa que consiste em padrões de permissão, geração, prevenção, produção e estímulo.

A organização ativa também distingue modelos mecanísticos de esquemas taxonômicos ou de sequências temporais. Existem modelos para explicar por que certos elementos são possíveis e impossíveis e que permitem até mesmo prever a existência de elementos não descobertos, mas a mera classificação taxonômica na tabela periódica apenas reúne elementos de acordo com suas estruturas atômicas; seus componentes não exibem qualquer atividade. Da mesma forma, modelos unicamente sequenciais, como descrições dos estágios do desenvolvimento de um organismo, apenas mostram como as partes estão organizadas em cada estágio, mas não mostram como uma configuração das partes é produzida a partir da configuração anterior.

A organização ativa nos mecanismos é sustentada por uma organização espacial e temporal das partes que os compõem. O que se entende por organização espacial são, por exemplo, tamanhos, formas, localizações orientações, direções, conexões e compartimentos dos componentes. Organização temporal refere-se à ordem, frequência e duração, por exemplo, das diferentes atividades. É possível que as mesmas entidades e atividades organizadas de formas diferentes produzam mecanismos distintos.

A análise funcional de Cummins é incompleta, nesse sentido, pois não leva em conta todos esses fatores. Ela exige apenas que a organização seja algo representado em um sistema; como um diagrama de caixas e setas. Isso é muito pouco, pois quase tudo pode ser descrito dessa forma: argumentos, sistemas taxonômicos e linhas de comando são exemplos. Ao contrário, explicações mecanísticas são corpóreas, no sentido que são ancoradas em componentes que ocupam espaço e tempo e engajam-se em atividades que levam tempo para ocorrer. Uma descrição mecanística não é meramente uma somatória das partes ou capacidades, mas uma descrição de como elas trabalham juntas.

2.4.7 Relevância constitutiva

Diante do que foi discutido até o momento, um ponto ainda persiste obscuro: necessário oferecer uma abordagem de relevância explicativa constitutiva, isto é, de quando uma parte de S é um mecanismo da ψ -ação de S. Nem todas as partes de S são mecanismos. Algumas fazem parte de S sem, no entanto, participar de qualquer mecanismo.

Para Craver, uma parte de um componente é um mecanismo se é possível mudar o

comportamento do mecanismo como um todo pela manipulação da parte e mudar o comportamento da parte ao intervir no mecanismo como um todo. Isso é o que ele chama de abordagem da *manipulabilidade mútua*.

Formalmente, isso pode ser colocado da seguinte maneira: (i) X é parte de S ; (ii) nas condições relevantes para o interesse explicativo, existe uma mudança da φ -ação de X que muda a ψ -ação de S ; (iii) nas condições relevantes para o interesse explicativo, existe uma mudança da ψ -ação de S que muda a φ -ação de X .

O primeiro condicional acima pode ser colocado da seguinte forma:

(CR₁) Se φ é colocado no valor φ_1 em uma intervenção ideal, então ψ toma o valor $f(\varphi_1)$.

Isso reflete a importância de experimentos de baixo para cima. É importante observar que CR₁ não se refere a condições quaisquer, mas sim as condições que são explicativamente relevantes. Isso precisa ser julgado caso a caso. Quando se quer explicar como um organismo funciona quando está saudável, tais condições não podem ser patológicas ou condições laboratoriais. Da mesma maneira, quando se quer saber como um organismo funciona sob condições patológicas ou industriais, essas devem ser as condições levadas em questão.

A exigência de CR₁ também é capaz de excluir os efeitos estéreis e meros correlatos. Ainda assim, CR₁ não é necessário nem suficiente para a φ -ação de X ser relevante para a ψ -ação de S . É possível, por exemplo, que ocorram respostas compensatórias de tal forma que não haja mudança na ψ -ação de S , ainda que X seja um componente relevante de S . Além disso, intervenções em condições de fundo podem mudar a ψ -ação de S , embora não façam parte do mecanismo. Craver admite que a distinção entre condições de base e componentes reais é feita com base em interesses pragmáticos. Ainda assim, ele defende que isso pode ser feito objetivamente, usando experimentos de interferência e estimulação conjuntamente, ou experimentos de ativação. Além disso, condições de fundo são inespecíficas, afetando muitos fenômenos além daquele em estudo.

O segundo condicional da abordagem de manipulabilidade mútua pode ser simplificado da seguinte maneira:

(CR₂) Se ψ é colocado no valor ψ_1 em uma intervenção ideal, então φ adquire o valor $F(\psi_1)$.

CR₂ descreve os efeitos de experimentos de cima para baixo. Craver recorda que é preciso levar em consideração os efeitos de mera correlação e ativação tônica, discutidos acima. Isso pode ser contornado usando os dois condicionais associados.

Relevância etiológica (causal) e relevância constitutiva são conceitos distintos. A relação de relevância constitutiva é sempre simétrica, ao contrário da relevância causal que, na maior parte das vezes, é unidirecional. Além disso, em relações constitutivas, existe uma dependência lógica entre uma instância de φ e uma instância de ψ , uma vez que a segunda é, em parte, constituída pela primeira. Além disso, como a relação constitutiva é sincrônica, a aquisição de determinado valor por φ não precede a tomada de determinado valor por ψ . Por esse motivo, não se deve pensar nas relações internível como causais.

Diante disso, Craver propõe um esboço do modelo de relevância constitutiva. Uma intervenção ideal em φ com relação a ψ é uma mudança em φ que muda o valor de ψ somente através da mudança em φ . Isso implica que:

(I1_c) a intervenção em I não muda ψ diretamente;

(I2_c) I não muda o valor de uma outra variável φ^* que muda o valor de ψ a não ser através da mudança em φ ;

(I3_c) I não está correlacionado a uma outra variável M que é causalmente independente de I e também uma causa de ψ ;

(I4_c) I fixa o valor de φ de tal forma que exclui da contribuição de outras causas de φ ao valor de φ .

Para Craver, relações de manipulabilidade mútua podem e devem substituir o critério de derivabilidade como um ideal regulador em explicações constitutivas em neurociência. Não é necessário derivar um fenômeno da descrição de um mecanismo, mas sim mostrar como ele se situa na estrutura causal do mundo. É preciso saber como o fenômeno muda sob uma série de intervenções nas partes e como as partes mudam quando alguém intervém para mudar o fenômeno. Isso permite fazer previsões sobre como o sistema se comporta sob uma variedade de condições, o que abre portas para intervir e modificar o sistema nervoso, satisfazendo assim os objetivos explicativos da neurociência.

2.5 Níveis

Explicações em neurociência abrangem múltiplos níveis. O termo “nível” é ambíguo e possui vários usos comuns em neurociência contemporânea. Aqui, usamos a concepção que Craver chama de *níveis de mecanismos*.

As várias acepções de níveis diferenciam-se em alguns aspectos. Em primeiro lugar, elas se

referem a diferentes tipos de objetos que são distribuídos nos níveis: 1) produtos ou unidades da ciência; 2) objetos do mundo, como atividades, entidades, propriedades e estados.

Além disso, deve-se levar em conta a relação internível em questão. No caso dos níveis de ciência, a relação pode ser entre: 1) os produtos da ciência (análises, descrições, modelos explicativos ou teorias); 2) unidades da ciência (campos, paradigmas, perspectivas e programas de pesquisa). No caso dos níveis da natureza, ou objetos do mundo, a relação pode ser de: 1) causalidade; 2) tamanho; 3) composição.

Por fim, a última diferença é aquilo que faz com que dois itens estejam no mesmo nível: o tamanho, as forças a que estão sujeitos, as leis que os governam e a frequência de interação entre eles. Os níveis de mecanismos, segundo essa concepção de Craver são níveis da natureza com relações de composição entre eles.

A seguir, falaremos brevemente de cada concepção de nível, situando os níveis de mecanismos nesse contexto.

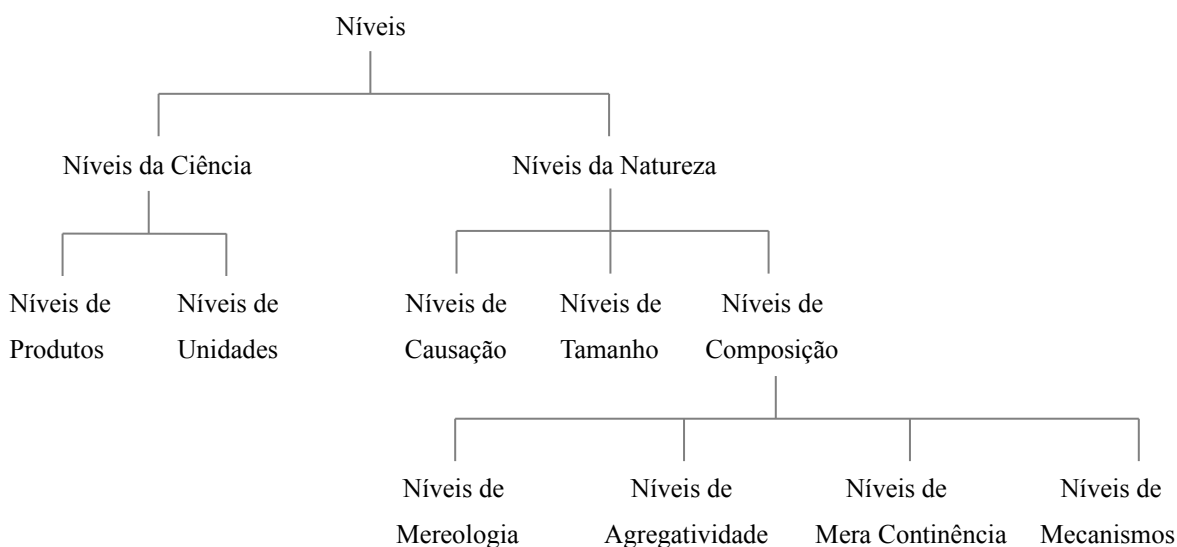


Figura 1. Uma taxonomia de níveis

2.5.1 Níveis da ciência

Frequentemente, na literatura filosófica, níveis são definidos com base em divisões da ciência, e não da estrutura do mundo. Oppenheim e Putnam (1958) defendem que as unidades e produtos da ciência correspondem um ao outro e também a divisões na estrutura do mundo. Eles definem seis níveis “mereológicos” da natureza: partículas elementares, átomos, moléculas, células, organismos e sociedades. Cada nível corresponde a um vocabulário teórico, a um conjunto de leis e a uma ciência. A unidade da ciência é atingida explicando o fenômeno no domínio de um campo da

ciência com teorias de outro campo da ciência.

Craver julga que essa visão é demasiado simplista. Além disso, ele critica a suposta correspondência entre níveis da natureza, níveis de unidades e níveis de produtos da ciência. Na sua concepção, as teorias são compostas por itens pertencentes a vários níveis “mereológicos”. Estes são integrados em uma só teoria não porque podem ser descritos pelas mesmas leis, mas porque a teoria é capaz de descrever, prever, explicar e testar a ocorrência de fenômenos. Além disso, várias teorias podem coexistir em um mesmo nível, a depender do domínio de pesquisa do cientista. Por esse motivo, não há correspondência entre níveis da ciência e níveis da natureza.

2.5.2 Níveis da natureza

Níveis da natureza compreendem objetos como entidades, atividades, propriedades e mecanismos. Relações entre níveis da natureza são de causação, tamanho e composição. Na concepção de Craver, essas relações devem ser diferenciadas umas das outras a fim de evitar as ambiguidades resultantes da sua compreensão incompleta.

Níveis causais, de controle ou de processamento são estágios de uma tarefa ou processo. Nesse sentido, os estágios se relacionam uns aos outros de acordo com uma sequência causal e, portanto, temporal. Exemplos de níveis de processamento em neurociência são as denominações de áreas cerebrais como primárias ou secundárias, como no caso dos córtex somatossensoriais primário e secundário e os córtex auditivos primário e secundário.

A principal diferença entre explicações em níveis de processamento e explicações em níveis mecanísticos é que, enquanto a primeira é constituída por relações entre itens distintos, a segunda é formada por relações entre parte e todo. Além disso, os níveis de processamento organizam-se temporal e sequencialmente, enquanto os níveis mecanísticos ocorrem no mesmo tempo do fenômeno. Por fim, níveis mecanísticos não possuem as relações causais características dos níveis de processamento, uma vez que atividades de níveis inferiores são estágios de atividades de níveis superiores, e não causas delas.

Níveis de controle são ligeiramente diferentes de níveis de processamento. Seus objetos são agentes relacionados por subordinação, ou seja, níveis superiores controlam as atividades dos níveis inferiores. É nesse sentido que se diz que genes controlam o desenvolvimento e que o córtex pré-frontal controla outras regiões do cérebro e o comportamento. Diferentemente dos níveis mecanísticos, nos níveis de controle as hierarquias mais altas e as mais baixas são partes distintas de um sistema maior.

Os *níveis de tamanho* são a descrição mais comum dos níveis de natureza. Aqui, os objetos

são entidades classificadas de acordo com seu tamanho relativo. Segundo Wimsatt, que compartilha alguns pontos com essa noção, níveis aparecem como picos de regularidade e previsibilidade em uma escala de tamanho aproximadamente logarítmica. Objetos do mesmo nível tendem a interagir mais entre si do que com objetos de níveis diferentes.

Essa classificação apresenta vários problemas. Em primeiro lugar, é necessário esclarecer se (i) há interações causais entre objetos de níveis diferentes e, caso haja, (ii) se essas relações são menos comuns que as existentes entre objetos do mesmo nível. Craver concorda com Wimsatt que interações causais entre objetos de tamanhos diferentes ocorrem. Objetos grandes, como planetas, podem interagir com objetos pequenos, como moléculas. Entretanto, quando se admite que níveis de tamanho possuem uma relação composicional, a resposta não é tão evidente. O conceito habitual de causação está impregnado de noções implícitas, como a independência lógica entre causas e efeitos e a precedência de causas sobre efeitos, que são muito difíceis de conciliar com a ideia de relações composicionais. Quanto à segunda questão, Craver se abstém de oferecer uma resposta.

Além das questões a respeito de interações entre níveis, há também as que se referem às regularidades. Craver indaga se (iii) há regularidades entre objetos de níveis diferentes e, caso positivo, (iv) se regularidades entre objetos situados no mesmo nível são mais comuns que aquelas com objetos de níveis diferentes. Sua posição é de ignorância nesse sentido. Embora admita prontamente que existam regularidades entre objetos de níveis diferentes, ele prefere não se posicionar quanto à questão (iv).

O principal problema em compreender níveis mecanísticos como níveis de tamanho é que as relações de tamanho entre níveis mecanísticos são apenas um achado incidental de uma relação *composicional* entre esses níveis. O que realmente importa não é que certos objetos sejam menores, ou maiores que outros, mas que eles possuam ou não algum tipo de relação parte-todo.

Os *níveis de composição* propostos por Craver merecem uma discussão mais aprofundada. Craver define quatro tipos de composição: mereológica, agregativa, espacial/material e mecanística. O último é o mais importante para as explicações multiníveis em neurociência.

2.5.2.1 Níveis de mereologia

A relação mereológica, ou de parte-todo, é a relação composicional mais comum nas discussões sobre níveis. Existem vários sistemas mereológicos. A maioria deles trata qualquer objeto complexo como um agregado, ignorando as relações entre as partes. Além disso, Craver argumenta que muitas das premissas dos sistemas mereológicos são equivocadas. Por exemplo, o *teorema da reflexividade* afirma que todo objeto é parte de si mesmo. Para ele, isso implicaria em

dizer que todos os níveis em neurociência estão em um nível mais alto e mais baixo que ele mesmo. Isso é problemático porque, nas explicações neurocientíficas, os níveis são exclusivos, ou seja, cada objeto aparece em uma única hierarquia. Além disso, essa noção é incompatível com uma visão ôntica da explicação como a de Craver, pois significaria dizer que cada objeto é uma explicação de si mesmo.

Além disso, a maioria das abordagens da mereologia admitem o *teorema da extensionalidade*, segundo o qual cada objeto é completamente determinado pelo conjunto das suas partes. Isso não leva em conta um aspecto muito importante dos mecanismos em neurociência, ou seja, de que as partes de níveis mais baixo são *organizadas* nos todos que ela compõem.

Por fim, Craver afirma que as abordagens mereológicas podem ser usadas para tratar de objetos abstratos e concretos. No entanto, em níveis de explicações neurocientíficas, é importante resguardar as propriedades material e espacial das partes.

2.5.2.2 Níveis de agregatividade

Nos níveis de agregatividade, os objetos são as propriedades dos todos e a propriedades das partes. Assim como os níveis de mereologia, os níveis de agregatividade se revelam inadequados para tratar da neurociência, uma vez que não levam em conta a organização das partes e suas propriedades.

2.5.2.3 Níveis de mera continência material/espacial

Nesse sistema, a demarcação dos níveis dos objetos é feita pelas barreiras espaciais que os contém. Segundo essa concepção, uma entidade está em um nível mais baixo que outra se a entidade de menor nível está nas fronteiras espaciais da entidade de maior nível e compõe uma parte da sua massa. O problema dessa abordagem é sua incapacidade de distinguir meros pedaços do sistema e reais componentes do sistema. Por essa razão, ela é inadequada para as explicações em neurociência.

2.5.2.4 Níveis de mecanismos

Na concepção de Craver, níveis de mecanismos são níveis de composição, mas as relações existentes não são simplesmente espaciais ou materiais. Aqui, os objetos são, nos níveis mais altos, mecanismos que se comportam de alguma forma e, nos níveis mais baixos, os componentes desses

mecanismo. Eles não devem ser concebidos como entidades, nem atividades, mas sim como entidades que agem. Os componentes de nível mais baixo são *organizados* para formar componentes de nível mais alto. Nesse sentido, as eles se relacionam espacial, temporal e ativamente uns com os outros, ou seja, trabalham juntos para fazer algo.

Os níveis de mecanismos possuem alguns aspectos dos níveis de Wimsatt, descritos acima. Eles são componenciais, os componentes são tipicamente maiores ou, no máximo, do mesmo tamanho do mecanismo como um todo e são locais de generalizações estáveis e, portanto, de regularidade e previsibilidade. Entretanto, tais características não são suficientes no caso dos níveis de mecanismos. Eles não devem ser concebidos como níveis de meros objetos, mas sim de componentes ativos. Muitas vezes, como no caso da sinapse neuronal, os componentes não correspondem a entidades com limites bem definidos. O que caracteriza cada componente como tal é o seu comportamento organizado.

Os níveis de mecanismos também não são divisões monolíticas da estrutura do mundo. Embora o mundo exiba regularidade e previsibilidade dentro de algumas divisões, como átomos, moléculas e células, os níveis de mecanismos vão além dos que essa divisão sugere. A hierarquização dos componentes só pode ser feita caso a caso. Cada componente pode estar em um nível diferente, a depender do mecanismo do qual está participando. Não existem tipos objetivos de componentes, independentes da sua organização. Assim, só é possível hierarquizar os componentes à medida que são feitas novas descobertas sobre a relevância de cada um deles dentro de um mecanismo.

2.6 Explicação não fundamentalista

Existem argumentos metafísicos que rejeitam a possibilidade de propriedades não fundamentais terem poderes causais sobre fenômenos fundamentais. Esse ponto de vista, chamado fisicalismo metafísico, opõe-se à existência de explicações não fundamentais. Nos próximos itens, examinaremos alguns pontos pertinentes das explicações não fundamentais e de como ela pode ser sustentada diante dos argumentos fisicalistas.

2.6.1 Relevância causal e fazer diferença

Uma variável X é causalmente relevante para uma variável Y se, e somente se há uma intervenção ideal em X que muda o valor de Y ou a probabilidade de distribuição de valores de Y . Se as causas são os fatores que fazem diferença para o fenômeno, então a noção de relevância

causal deve dar atenção apenas a eles.

Como já foi discutido antes, a abordagem de Craver é bastante diferente de outras abordagens causais. Suas generalizações não demandam que a relação entre X e Y seja universal, nem que o fenômeno seja subsumido a leis naturais, podendo valer apenas para condições relevantes para o pesquisador, por exemplo. Elas também são capazes de dar conta de fenômenos estocásticos, que se comportam de forma irregular, bem como de fenômenos mecanisticamente frágeis. Por fim, elas são capazes de diferenciar causas verdadeiras de meros correlatos, efeitos de causa comum e detalhes irrelevantes.

Como já dissemos antes, para explicar um fenômeno, não é necessário nem suficiente haja uma lei da natureza entre a variável e o *explanandum*, mas sim que esse fenômeno seja situado na estrutura causal do mundo. A explicação tampouco exige a transmissão de marcas ou a existência de algum tipo de conexão física entre os eventos, à maneira de Salmon e Dowe.

Finalmente, essa abordagem revela-se capaz de lidar com o problema de causação por omissão e prevenção, uma vez que usa conceitos de relevância causal, e não poder causal. Ausências não são objetos e, portanto, não tem propriedades e não podem exercer poderes causais. Contudo, elas podem fazer diferença para ocorrência do fenômeno e, nesse sentido, são causalmente relevantes.

2.6.2 Contrastes e pontos de gatilho

De acordo com a abordagem manipulacionista, uma variável X é relevante para uma variável Y somente se houver uma intervenção ideal em X que muda o valor de Y apenas através da mudança em X. Isso não quer dizer que qualquer mudança em X seja capaz de mudar o valor de Y. A formulação de contrastes do tipo “Sócrates ingerir cicuta e não uma outra bebida qualquer causou sua morte” permite lidar com essa particularidade.

Craver ressalta a importância da formulação de contrastes para se falar sobre relevância causal e explicativa. O primeiro motivo é que, para dizer se uma alegação causal é verdadeira ou falsa, é necessário explicitar os contrastes na descrição da causa e na descrição do efeito. Retornando ao exemplo de Sócrates, Craver defende que afirmações do tipo “Sócrates ingerir cicuta causou sua morte” é ambígua. Ela pode se referir ao contraste entre ingerir e bebericar cicuta, que não fazem diferença para a sua morte. Também podem se referir ao contraste entre ingerir cicuta e água, o que, por sua vez, é relevante para o desfecho do fenômeno.

Da mesma forma, formulações contrastivas também lidam com ambiguidades nas alegações causais quantitativas. Dizer, por exemplo, que fumar um maço de cigarro por dia causa câncer

depende do contraste a que se refere. Alguém que reduz o consumo de cigarro de dois para um maço por dia diminui sua chance de ter câncer de pulmão, mas alguém que aumenta de meio para um maço por dia aumenta essa chance. Embora a circunstância seja a mesma, fumar um maço de cigarro por dia, os contrastes geram diferentes alegações causais.

A importância das formulações contrastivas se faz ver nos próprios experimentos conduzidos para testar hipóteses neurocientíficas. Tais experimentos envolvem a comparação do valor de uma variável sob condições experimentais com o valor da variável sob condições de controle. Um experimento é bem conduzido quando o grupo experimental e o grupo controle diferem apenas em relação ao valor da variável causal putativa.

Para decidir entre contrastes causais apropriados, é necessário buscar *pontos de mudança*, nos possíveis valores da variável causal que correspondam às mudanças dos valores da variável de efeito. Pode haver apenas um ponto de mudança, mas alguns casos mais complexos podem envolver dois ou mais pontos de contraste.

2.6.3 Poderes causais em mecanismos de alto nível

À primeira vista, o fisicalismo metafísico não admite que um mecanismo possua qualquer relevância causal, pois toda a causalidade se concentra nas suas partes. Craver aponta, contudo, que existem duas questões distintas: primeiro, a questão metafísica da existência de propriedades com poderes causais; segundo, a questão experimental da relevância causal, ou seja, se existem generalizações do tipo “em condições W, a intervenção para mudar o valor de X de x_1 para x_2 muda o valor de Y para $f(x_2)$ ”. Apenas a segunda se coloca no presente trabalho. É importante ressaltar que esse tipo de indagação se refere à relevância causal e explicação causal, não a poderes causais.

Ainda que a noção de que a relevância causal está fundamentada em poderes causais seja atrativa, ela enfrenta três desafios metafísicos. Em primeiro lugar, é necessária uma abordagem de poderes causais esclarecedora, o que requer explicar por que relações de relevância causal se sustentam e por que eles satisfazem normas como E1 a E5. O segundo é fornecer uma abordagem convincente de causação por omissão e prevenção. Por fim, é necessário fornecer uma abordagem de poderes causais que faça sentido no contexto de práticas experimentais e observacionais que os cientistas usam para descobrir relações causais. Dados esses desafios, Craver sugere que a questão metafísica seja rejeitada em favor de uma visão em que a causação refira-se a fazer diferença, independente de isso ser feito através do exercício de atividades ou poderes.

2.7 A unidade mosaiforme da neurociência

A neurociência é um programa de pesquisa que engloba vários campos do conhecimento, como anatomia, bioquímica, ciência da computação, radiologia, biologia molecular, eletrofisiologia, psicologia experimental, etologia, farmacologia e psiquiatria, por exemplo. Essa heterogeneidade torna necessária uma abordagem da unidade da neurociência que seja coerente com a prática dos cientistas de suas diversas áreas.

Seguindo uma visão tradicional, a unidade da neurociência seria obtida através de modelos de redução interteórica. Existem vários argumentos em favor desse ponto de vista. Em primeiro lugar, a redução interteórica pode ser definida precisamente através da lógica formal, permitindo uma avaliação precisa. Em segundo lugar, devido à longa tradição de uso de modelos reducionistas na filosofia da física, da química e da biologia, é razoável esperar que o mesmo ocorra nas neurociências. Por último, desde o manifesto de Oppenheim e Putnam (1958), redução tem sido praticamente um sinônimo da unidade explicativa da ciência.

Contudo, todos os sistemas de redução possuem problemas que já foram discutidos anteriormente. A proposta de Craver é mostrar que a neurociência pode ser unificada não pela redução de todos os fenômenos ao nível fundamental, mas usando resultados de campos diferentes para adicionar restrições uma explicação mecanística multinível. Como cada campo aborda o problema sob uma perspectiva diferente, usando diferentes pressupostos e técnicas, as evidências que eles propiciam tornam as explicações robustas. Essa unidade é um objetivo para a neurociência não porque permite distinguir a ciência da não-ciência, nem por que mostra que tudo está ligado ao nível fundamental, mas sim porque tal unidade é um sinal de sucesso epistêmico.

2.7.1 A redução na história da neurociência

Mesmo alguns defensores das teorias reducionistas, como Schaffner e Churchland, admitem que esse modelo não descreve de forma adequada a história e a prática da neurociência. Para Craver, abordagens que ignoram as práticas dos neurocientistas não são adequadas. A neurociência não possui os mesmos objetivos explicativos da física, já que as generalizações em neurociência são mais frágeis, variáveis e historicamente contingentes do que as leis físicas. Ademais, os objetos no domínio da neurociência são multiplamente realizados, de modo que encontrar uma redução clássica não deve ser considerado uma prioridade para a neurociência. Em neurociência, tais reduções só se dariam, por exemplo, em membros normais de um subgrupo de uma espécie em uma época particular do desenvolvimento.

Por outro lado, outros defensores da redução argumentam que ela descreve adequadamente a prática científica, uma vez que hoje é possível formular explicações para fenômenos de alto nível em termos de mecanismos moleculares. Em contrapartida, Craver sustenta que os avanços em biologia molecular não mostram que as melhores explicações em neurociência estão no nível molecular. Para compreender as normas da explicação neurocientífica, é necessário olhar para o que os cientistas de fato fazem. Para isso, Craver faz uma análise histórica da descoberta da sensitização de longo prazo (SLP), um dos mecanismos implicados na formação da memória, e que comumente é usado como exemplo de redução interteórica.

2.7.2 As origens da SLP

A maioria dos neurocientistas remonta a descoberta da SLP a uma série de artigos publicados por Tim Bliss, Terje Lømo e Tony Gardner Medwin, em 1973. Contudo, Craver mostra que essa descoberta foi precedida por uma série de eventos que culminaram nessas publicações.

A primeira publicação mostrando a produção da plasticidade sináptica no hipocampo data da década de 50, em estudos *in vivo*. Nessa época, esse evento era considerado apenas um instrumento laboratorial, de pouca importância e sem relação com o aprendizado ou memória. Na realidade, os neurocientistas da época não localizavam a memória em uma região específica do cérebro, mas acreditavam se tratar de uma função global do cérebro. O hipocampo, hoje considerado uma das estruturas cerebrais mais envolvidas com a formação da memória, era associado a outras funções, como olfato, inibição comportamental, emoção, medo, apetite, atividade sexual, sono e respiração.

Na realidade, as pesquisas dirigidas ao hipocampo tinham outros motivos que não a investigação da memória. Em primeiro lugar, o hipocampo tem relação com a epilepsia, o que já era conhecido à época. Além disso, ele é um valioso modelo experimental para o estudo da eletrofisiologia dos circuitos neurais. Por fim, pelo fato de o hipocampo de roedores ser grande e acessível através do crânio, ele era uma região atraente para estudos *in vivo*.

O hipocampo era considerado um modelo experimental útil para integrar resultados da anatomia e da eletrofisiologia. Contudo, essa integração não ocorria entre níveis, mas em um único nível através da adição sucessiva de restrições em um mecanismo neural. Por exemplo, alguns cientistas faziam pesquisas histológicas com coloração de Golgi que mostravam padrões de conexão na região cerebral, enquanto outros usavam estudos de degeneração terminal que mostravam os axônios degenerados, construindo um diagrama anatômico que pudesse ser usado para investigação eletrofisiológica. Estudando tais diagramas, eletrofisiologistas podiam planejar experimentos de intervenção nas propriedades eletrofisiológicas de algumas células, verificar os efeitos de tais

intervenções em outras células e estudar a propagação da excitação neuronal. Os achados de eletrofisiologias e anatomistas não se relacionavam uns aos outros por redução, mas combinavam diferentes técnicas para avaliar aspectos de um mecanismo em um mesmo nível oferecendo evidências independentes sobre um único fenômeno.

2.7.3 A virada mecanística

Inicialmente, os cientistas tinham uma visão reducionista das mudanças sinápticas fundamento do aprendizado. Eles não argumentavam que o aprendizado seria idêntico à SLP, mas que mudanças sinápticas seriam tipos simples e primitivos de aprendizado. Com a publicação dos experimentos de 1973, começou a haver uma mudança do pensamento. Esses cientistas passaram a descrever a SLP não mais como um tipo de memória ou como idêntico à memória, mas como um componente do mecanismo multinivelado da memória.

Essa mudança de uma perspectiva reducionista para uma perspectiva mecanística estabeleceu duas metas para as pesquisas subsequentes. Em primeiro lugar, voltando-se para níveis inferiores, descobrir os mecanismos responsáveis pela SLP; em segundo lugar, com enfoque nos níveis superiores, avaliar o papel da SLP em mecanismos de memória de nível mais alto. Essa virada na perspectiva explicativa da SLP situou o fenômeno em uma estrutura teórica para integrar vários campos de pesquisa, como anatomia, bioquímica, eletrofisiologia, psicologia e psiquiatria.

2.7.4 O mecanismo como uma proposta de trabalho

A história da descoberta da SLP mostra que o programa de pesquisa em neurociência não é reducionista. Ao contrário, ele tenta descobrir conexões entre níveis mais altos e mais baixos em uma hierarquia de mecanismos. Embora muitos pesquisadores usem técnicas de biologia molecular, igualmente importantes são as pesquisas que relacionam a SLP a outros fenômenos de alto nível, como diferentes tipos de memória. Além disso, o foco dos modelos reducionistas nas relações entre níveis restringe a atenção que deveria ser dada à integração entre campos de um mesmo nível. Por fim, os pesquisadores no programa de pesquisa do SLP substituíram o reducionismo pela pesquisa de mecanismos multinivulares como um objetivo explicativo.

2.7.5 A integração internível e a unidade de mosaico da neurociência

Craver propõe um modelo positivo da unidade da neurociência. Para ele, os modelos

reducionistas não são capazes de lidar com casos de integração entre campos no mesmo nível, pois como os fenômenos não estão em níveis diferentes, os termos usados para descrevê-los não são traduzidos uns nos outros e as descrições não são homomórficas umas com as outras. Assim, o objetivo de pesquisa em neurociência deveria ser o de construir um mosaico em que múltiplos campos fazem contribuições pontuais para um esboço abstrato do mecanismo.

A construção de teorias mecanísticas ocorre através da acumulação gradual de conhecimentos pontuais no espaço dos mecanismos possíveis para um fenômeno. O *espaço dos mecanismos possíveis* contém todos os mecanismos que poderiam possivelmente explicar o fenômeno. As dimensões desse espaço são definidas pelas entidades, propriedades e atividades que compõem um mecanismo como-possivelmente em todos os níveis relevantes e pelas variedades de organização dentre os seus componentes. Nesse espaço, os pontos são explicações como-possivelmente. As distâncias entre os pontos indicam o grau de similaridade entre dois mecanismos possíveis e regiões desse espaço representam classes de mecanismos similares. Uma restrição é um achado que molda a forma das barreiras do espaço de mecanismos possíveis ou muda a sua probabilidade de distribuição sobre esse espaço. Campos diferentes da neurociência possuem diferentes problemas centrais, diferentes técnicas e diferentes vocabulários teóricos

Embora não seja possível saber de antemão quantos tipos de restrições existem nem quantas restrições são relevantes para o entendimento de um mecanismo, existem algumas restrições gerais que derivam da organização característica dos mecanismos discutida anteriormente.

Restrições de componentes refere-se à delimitação das entidades e atividades envolvidas em um mecanismo. O estoque de componentes expande-se e contrai-se através do tempo. Atualmente, sabe-se que o cérebro é composto de neurônios, que os neurônios transmitem sinais na forma de potenciais de ação, que eles se comunicam através de sinapses químicas, que são organizados em redes de células que formam sistemas, etc. Esse é o estoque ôntico da neurociência contemporânea.

As restrições de espaço compreendem três aspectos: a *organização* espacial, que se refere às barreiras bem definidas que separam os componentes ou os mecanismos, como a membrana nuclear e a membrana celular; à *localização* de diferentes componentes, entidades e atividades; e à *estrutura e orientação* do mecanismo.

As restrições temporais compreendem: a *ordem* das atividades e entidades que compõem um mecanismo, ou seja, a sua posição relativa nas séries, bifurcações e ciclos que constituem o mecanismo; as *taxas* e *durações* dos diferentes estágios do mecanismo.

Condições ativas são as relações entre as variáveis sob uma intervenção ideal e as condições sob as quais essas relações se dão. A integração entre campos se dá através da intervenção em uma variável para detectar a outra, usando técnicas de diferentes campos.

2.7.6 Redução *versus* integração de campos no mesmo nível

Sob a perspectiva da abordagem MMM, a unidade da ciência é obtida através da integração de resultados de diferentes campos de pesquisa, no mesmo nível mecanístico. Esse tipo de integração não pode ser explicado pelas abordagens reducionistas, pois os termos teriam que estar em níveis diferentes para que fossem traduzidos de um campo para o outro. Além disso, as restrições não são identificadas umas com as outras, nem homomórficas umas com as outras. Esse tipo de integração internível não é levado em conta pelas abordagens reducionistas da unidade da neurociência.

Quando neurocientistas explicam um fenômeno, seu objetivo não é construir uma teoria, mas sim identificar um mecanismo, seus componentes e as características cruciais de sua organização espacial, temporal e ativa. A unidade de mosaico da neurociência é obtida à medida que pesquisadores em campos diferentes integram seus resultados como restrições a explicações mecanísticas multinivulares.

O modelo de mosaico possui muitas vantagens sobre a redução para compreender as formas de integração entre campos no mesmo nível. Em primeiro lugar, ele fornece uma forma direta de interpretar “níveis” e, analogamente, a idéia de integração internível. Em segundo lugar, enquanto os modelos reducionistas envolvem relações globais entre teorias em diferentes níveis, o modelo de mosaico é compatível com a prática neurocientífica, em que relações internível são frequentemente formuladas passo a passo, dentro de mecanismos locais, através da adição de restrições em relações internível. Em terceiro lugar, a abordagem mecanística acomoda tanto integração internível para cima como para baixo. Por fim, a visão de mosaico detalha as várias forças que movem a coevolução do trabalho em níveis diferentes.

Tomando-se o exemplo do mecanismo de formação da memória (anexo 4), pode-se distinguir pelo menos quatro níveis: a memória e o aprendizado, no nível mais alto, a geração de mapas espaciais no hipocampo, as atividades químicas e elétricas das sinapses e células e, no nível mais baixo, as atividades das moléculas que compõem a sinapse. Integração internível para cima consiste em mostrar que um item é um componente de um mecanismo de nível mais baixo. Integração internível para baixo consiste em descrever mecanismos de nível mais baixo para um mecanismo de nível mais alto. Diz-se que um fenômeno está integrado com seus níveis adjacentes quando se reconhece: 1) quais os aspectos relevantes do fenômeno no nível em que ocorre; 2) como esses aspectos se organizam com um mecanismo de nível mais alto; 3) como esses aspectos do fenômeno são constitutivamente explicados por um mecanismo de nível mais baixo. Essa visão de integração internível fornece uma abordagem muito mais precisa e cientificamente informada do

que é requerido para ligar níveis em uma hierarquia de mecanismo do que os modelos reducionistas. Integrar níveis mecânicos não é o mesmo que estabelecer identidades entre níveis, mas sim estabelecer relações de componência e relevância explicativa.

Assim como a integração entre campos de níveis diferentes, a integração entre campos de mesmo nível é construída através da descoberta passo a passo de restrições em um mecanismo multinível. Um dos objetivos das ciências que buscam mecanismos é acomodar a taxonomia do fenômeno explanandum à taxonomia das explicações mecânicas. Essa busca leva cientistas de diferentes campos a mudar a forma como eles pensam sobre seus domínios a fim de construir pontes com outros campos entre níveis. Essa acomodação pode ocorrer tanto de baixo para cima como de cima para baixo. Por exemplo, cientistas de níveis mais altos influenciam aqueles de níveis mais baixos através do estabelecimento de barreiras para o mecanismo. Da mesma forma, níveis mais altos podem precisar ser reformulados para acomodar as descobertas de níveis mais baixos. Isso é o que Craver chama de *restrições de acomodação*.

Restrições espaciais compreendem a localização dos diferentes componentes do mecanismo, assim como a localização do próprio mecanismo. Restrições temporais também podem ser importantes para distinguir fenômenos semelhantes, como é o caso da memória de curto e longo prazo.

Restrições de manipulabilidade são obtidas através de experimentos multinível feitos por pesquisadores de campos diferentes, mostrando como fenômenos em níveis diferentes são mutuamente manipuláveis.

Os campos da neurociência são frequentemente integrados quando pesquisadores de diferentes campos encontram restrições de níveis diferentes em um mecanismo multinível ou colaboram para ligar os níveis em um mecanismo. Essa forma de integração é muito diferente da tentativa de traduzir a teoria de um campo na teoria de outro campo ou de criar imagens homomórficas de um fenômeno estudado em campos diferentes. Na realidade, neurocientistas começam com um objetivo inespecífico, como explicar a memória espacial e, sucessivamente, descobrem restrições que permitem acomodar modelos como-possivelmente até chegar a um modelo como-verdadeiramente. Tais restrições são frutos dos experimentos de vários cientistas de campos diferentes que investigam aspectos diferentes do mecanismo. As restrições são usadas uma a uma para a construção de um esboço de um mecanismo, mostrando passo a passo quais os componentes e propriedades são relevantes para o mecanismo e como eles se organizam espacial, temporal e ativamente no mecanismo.

O modelo mosaiforme da unidade da ciência é valioso não por servir como um critério de distinção entre ciência e não-ciência, nem por permitir relacionar todos os fenômenos a um nível

fundamental, mas porque tem a virtude epistêmica de produzir explicações robustas que podem se sustentar diante o exame minucioso a partir de perspectivas diferentes, com técnicas diferentes. Os campos possuem perguntas, vocabulários teóricos, técnicas de investigação e domínios do fenômeno diferentes uns dos outros e, por esse motivo, campos diferentes fornecem maneiras independentes de acessar aspectos diferentes de um mecanismo. Uma descrição adequada de um mecanismo deve resistir ao exame dessas diferentes perspectivas. Ademais, experimentos realizados em níveis diferentes fornecem caminhos independentes para acessar fenômenos que são parte da mesma teoria multinívelar, além de situar o fenômeno em um corpo teórico que tanto explica o fenômeno como mostra o papel explicativo que o fenômeno possui.

Capítulo 3 – Críticas ao modelo mecanístico-causal multinível

A publicação do trabalho de Machamer et al, em 2000, foi um marco sobre a literatura em sobre explicações em neurociência. Nele, os autores expuseram os principais conceitos que culminaram na formulação do modelo mecanístico-causal. Foi o primeiro trabalho em que o conceito de mecanismo foi fornecido explicitamente: “mecanismos são entidades e atividades organizadas de tal forma que elas são produtivas de mudanças regulares das condições de início ou de instalação para as condições finais ou de terminação”. (Machamer et al, 2000, p. 3). Isso caracteriza o que eles chamam de “abordagem dual” de mecanismos, fundada nos conceitos de *entidades e atividades*.

Também nesse trabalho, os autores declaram seu comprometimento com uma existência ontológica tanto das atividades como das entidades. Mudanças de propriedades de entidades não são suficientes para descrever um mecanismo, uma vez que mecanismos *fazem coisas*. Essas coisas são as atividades, que são descritas em termos de “tipos de causas” (Idem, p. 6). Por esse motivo, uma descrição ontologicamente adequada de um mecanismo deve incluir tanto as entidades como as atividades, que possuem existência interdependente (não há atividades sem entidades e vice-versa). Ademais, os autores argumentam em favor de descrições multiníveis de mecanismos, nos quais as relações hierárquicas devem ser compreendidas como relações de parte-todo em que as entidades de nível mais baixo são componentes de mecanismos que resultam em entidades de nível mais alto, os fenômenos a serem descritos.

Até então, somente os aspectos mecanístico e multinível das explicações neurocientíficas haviam sido abordados. Em 2005, Craver expõe sua abordagem sobre a unidade da neurociência. Nesse trabalho, ele se coloca contra visões tradicionais que defendem a unidade da neurociência a partir de modelos de redução interteórica. Reconstruindo a história dos experimentos que levaram à descoberta da PLD, Craver argumenta que a descoberta de mecanismos é feita através da realização de experimentos internível que, frequentemente, requerem habilidades de diferentes áreas da ciência. A unidade da neurociência é obtida, portanto, quando múltiplos campos juntos delimitam um mecanismo que explica um fenômeno neurocientífico.

A obra de 2007 em que foi centrado o segundo capítulo do presente trabalho é, portanto, uma compilação das ideias em torno dessa abordagem de mecanismos, que batizamos de abordagem mecanístico-causal multinível, ou, simplesmente, MMM.

Desde o primeiro trabalho (Machamer et al, 2000), essas noções têm sido extensivamente criticadas e discutidas na literatura filosófica. São objetos de críticas, por exemplo, multinívelidade (Bickle, 2003, 2006), os mecanismos (Weber, 2008), e a questão da relevância

constitutiva das partes do mecanismo (Couch, 2011). Nesse capítulo, descreveremos com detalhes essas críticas para verificar até que ponto elas fragilizam o modelo MMM.

3.1. Bickle e sua abordagem “implacavelmente redutiva”

Em 2003, Bickle já havia publicado a sua proposta para a filosofia da neurociência, que ele batizou de “implacavelmente redutiva”. De modo geral, ele defende uma abordagem filosófica dessa disciplina que seja, ao mesmo tempo, parte dela. Baseando-se em um vasto conhecimento sobre as mais recentes descobertas em neurociência, Bickle defende que a maioria dos filósofos ignoram os conhecimentos recentes da área, posicionando-se normativamente de uma forma metafisicamente orientada.

Em 2006, ele volta a reforçar sua posição original. Segundo ele, embora impopular entre os filósofos, a posição reducionista é coerente com a própria forma como neurocientistas trabalham. O que ele chama de “redução”, aqui, é um pouco diferente da noção tradicional de redução interteórica em filosofia da ciência. Bickle se refere a ligações diretas mente-moléculas, ou seja, a explicação da cognição e do comportamento complexo a partir de neurociência de níveis inferiores, sem necessariamente apelar para definições lógicas formais. Para ele, os avanços obtidos em neurociência celular e molecular são tantos que “é possível reduzir conceitos e tipos psicológicos para vias e mecanismos moleculares” (Bickle, 2006, p. 412).

Um dos exemplos que Bickle usa para defender seu ponto de vista é um experimento de intervenção no nível genético do mecanismo molecular da formação da memória espacial, (anexo 2). Nesse experimento, cientistas provocam uma mutação da proteína CREB em camundongos, que impede a sua ativação. Isso resulta na indução de amnésia de longa duração para o reconhecimento social, mantendo intactos o aprendizado inicial e a memória social de curta duração. Para Bickle, isso é um exemplo de explicação reducionista em neurociência, na qual foi possível intervir em um nível mais baixo e medir uma mudança de comportamento estatisticamente significativa no nível do comportamento. Quando isso ocorre, níveis intermediários tornam-se supérfluos.

Bickle não afirma que se deva abandonar a pesquisa em níveis intermediários. Métodos como neuroimageamento funcional, modelamento de redes neurais, avaliações neuropsicológicas possuem papéis que ele considera indispensáveis, como nortear a criação de protocolos experimentais, direcionar os pontos de intervenção celular e molecular e medir e avaliar os tipos de atividades neurais envolvidas. Contudo, para ele, essas são todas questões heurísticas, que devem ser levadas em conta à medida que se busca os mecanismos moleculares e celulares que estão por baixo. Uma vez que os mecanismos de nível inferior sejam desvendados, a estratégia experimental

passaria a ser reducionista, intervindo no mecanismo de nível baixo e avaliando os efeitos disso sobre o comportamento.

Bickle deixa evidente que os avanços obtidos recentemente em neurobiologia celular e molecular precisam ser considerados e que uma abordagem explicativa que ignore os avanços dessa área não pode ser satisfatória. De fato, essas têm produzido resultados significativos muito rapidamente e, possivelmente, em uma velocidade que a neurociência cognitiva e os estudos de neuroimagem são incapazes de alcançar.

No entanto, apesar da importância dos estudos que ligam eventos moleculares diretamente ao comportamento, como o papel causal da proteína CREB na consolidação da memória, isso não é tudo o que há sobre a explicação dos fenômenos. Por exemplo, Eric Kandel, ganhador do prêmio Nobel e autor de um dos principais livros de neurociência (Kandel et al, 1997), apresenta dois capítulos sobre memória. No primeiro (cap. 35), ele descreve os processos da formação de memória em nível alto. Existem várias passagens que sugerem que Kandel considera causalmente relevantes para a formação da memória algumas partes macroscópicas do cérebro. Por exemplo, na página 520, ele diz: “durante a maior parte desse tempo, a opinião geral era a de que a memória seria uma propriedade geral do córtex cerebral como um todo. Na verdade, estudos recentes têm mostrado que a memória depende, realmente, de muitas regiões cerebrais”. Mais adiante, na página 523, ele diz: “Estudos em humanos e animais experimentais indicam que o sistema de memória no lobo temporal medial é bastante extenso. Esse sistema inclui o hipocampo, bem como o córtex entorrinal (...), o subículo (...) e os córtices para-hipocampais”.

Isso não quer dizer que Kandel adote uma explicação multinível por desconhecer os mecanismos moleculares da formação da memória. O capítulo 36 é inteiro dedicado a esse tema, incluindo o papel do AMP cíclico e da proteína CREB (pag. 536). Mesmo quando ele aborda as “surpreendentes possibilidades reducionistas sobre o enfoque neurobiológico do aprendizado” (p. 545), ele acrescenta: “obviamente, esses mecanismos celulares elementares estão embutidos em circuitos neurais dotados de poder computacional adicionalmente muito maior, o que pode somar um grau substancial de complexidade a esses mecanismos elementares”. Por fim, na conclusão do capítulo final do livro, Kandel reconhece abertamente a importância da multinívelidade:

Como tentamos mostrar neste livro, a fusão da biologia com a psicologia cognitiva é mais que o uso comum de métodos e de conceitos. A junção dessas duas disciplinas representa a convicção emergente de que as descrições científicas da mentação, *em diversos níveis distintos*, contribuirão eventualmente para uma compreensão biológica unificada do comportamento (Kandel, 1997, p. 552, grifo

nosso).

3.2. Schaffner e o papel das teorias, modelos e equações em neurobiologia

Como vimos, Schaffner (1993) defende que as explicações em biologia consistem em uma rede de sistemas aos quais se aplicam generalizações moldadas em uma terminologia multinível e multidisciplinar. As teorias em neurociência seriam “uma estrutura de modelos prototípicos interníveis, causais e temporais” (Schaffner, 2008, p. 1009), em que os modelos seriam formados por mecanismos prototípicos e suas variações. De acordo com ele, um modelo prototípico teria um escopo relativamente estreito de aplicação para poucos tipos. Um exemplo de mecanismo prototípico seria o potencial de ação que, com todas as suas particularidades, é aplicável apenas ao axônio da lula gigante, mas cujas variações permitem uma extensão para outros neurônios de outras espécies.

Até pouco mais da metade do século passado, houve tentativas de enquadrar as leis e teorias da biologia em uma estrutura lógica semelhante à da física através do uso de axiomatizações e equações diferenciais com ampla aplicação. Embora Schaffner (1993) reconheça o fracasso dessas tentativas para praticamente todos os exemplos em biologia, ele acredita que seria possível obter algo intermediário entre os mecanismos de estreito alcance de aplicação e os amplos modelos matemáticos. Esse seria o caso do modelo HH.

Nesse modelo, as leis que regem o circuito equivalente ao fluxo iônico através da membrana neuronal são leis físicas padrão, como a lei de Ohm e a equação de Nernst. Todavia, as equações que descrevem as correntes iônicas não são equações universais, mas sim equações empíricas construídas a partir dos dados obtidos pelo *clamp* de voltagem e escolhidas por sua simplicidade.

O modelo HH, segundo os próprios Hodgkin e Huxley (1952), possui aplicabilidade limitada a outros tecidos. Para eles, ainda que o mecanismo básico de condução possa ser o mesmo, as grandes diferenças nas formas dos potenciais sugerem que, para ser estendidas a outros casos, as equações precisariam de importantes adaptações em alguns dos parâmetros.

Além disso, eles reconhecem que as equações do modelo HH podem resultar de diferentes mecanismos moleculares. Ainda assim, o seu poder de previsão é muito grande, mesmo para mecanismos diferentes. Segundo Schaffner, isso revela o seu extremo grau de abstração, no sentido de poder suportar uma generalização muito ampla em termos das especificações das correntes, sem comprometimento com qualquer mecanismo molecular específico. De fato, o mecanismo de translocação iônica inicialmente proposto por Hodgkin e Huxley se revelou equivocado e hoje se sabe da existência uma enorme gama de canais iônicos, com estrutura e formas de ativação

diferentes.

Diante da diversidade de variações dos mecanismos para o potencial de ação, Schaffner (2008) propõe que equações com tal poder de generalização como as do modelo HH possam ter um papel unificador. Ele sugere que os pesquisadores devam buscar o que ele chama de “simplificações emergentes” desse tipo, situadas em um nível mais alto de abstração e que transcendem os detalhes do nível molecular. Um mecanismo abstrato, no qual os detalhes moleculares são ainda desconhecidos, podem fazer o papel de um mecanismo básico prototípico, que identifica as características centrais de uma entidade biológica e suas ações. Em alguns casos, essas características podem ser representadas por equações matemáticas que, embora se assemelhem em forma às equações físicas tradicionais, careçam de universalidade e exerçam sua função unificadora ao assumirem um papel prototípico.

Assim, segundo Schaffner, simplificações são essenciais para dar sentido aos processos neurais, tanto mais quanto mais complexo o sistema em análise. Contudo, essas simplificações devem estar direcionadas à produção de resultados em um nível emergente, mais abstrato, e não reducionista.

A noção de unificação proposta por Schaffner (2008) dá às equações em biologia um status justo. Embora não sejam, por si sós, explicativas, as equações obtêm sua importância pelo fato de oferecerem uma unificação de fenômenos. Não são, como argumenta Craver (2007) para as equações do modelo HH, “nada mais que uma descrição empírica do curso temporal das mudanças de permeabilidade” (Craver, 2007, p.54). De fato, como observa Schaffner, as equações possuem tal grau de abstração que admitem “generalizações muito amplas em termos da especificidade das correntes de membrana, mas não qualquer mecanismo molecular específico” (Schaffner, 2008, p. 1016). Ao suprimir detalhes sobre tais mecanismos, as equações tornam-se importantes instrumentos heurísticos para a pesquisa. Como observa Bogen (2008), elas admitem inferências das descobertas feitas com a lula gigante para axônios de outros animais. Mais ainda, as falhas de previsões pelas equações possibilitam a descoberta de diferentes mecanismos moleculares para o potencial de ação em diferentes espécies.

É interessante que o modelo de Hodgkin e Huxley atingiu esse nível de abstração antes mesmo dos conhecimentos recentes sobre os mecanismos subjacentes ao potencial de ação. Isso foi possível, segundo Schaffner, porque eles trabalharam em um nível de abstração acima dos mecanismos moleculares, o que ele chama de “simplificação emergente” (Schaffner, 2008, p. 1018). Parece-nos interessante e promissora a proposta de que um mecanismo mais abstrato como esse exerça o papel de um “mecanismo prototípico” (Schaffner, 2008, p. 1018) que identifica e caracteriza aspectos centrais de uma entidade biológica e suas ações.

3.3. Weber e as causas sem mecanismos

Marcel Weber (2005, 2008) defende que o modelo HH é um caso exemplar de explicação físico-química de um fenômeno biológico. Ele argumenta que, nesse caso, o potencial de ação é o *explanandum*, a descrição do mecanismo pode ser vista como as condições iniciais e as leis físico-químicas são as leis que regem essa ocorrência. “Em última análise”, prossegue ele, “é uma lei físico-química que explica por que os íons se movem através das membranas quando o potencial de ação se espalha” (Weber, 2005, p. 26).

Para Weber, o papel dos conceitos biológicos envolvidos (como neurônio, membrana) possuem um papel meramente descritivo, de delimitar a classe de sistemas e as condições que figuram em uma explicação puramente físico-química, como os diagramas de circuitos e as representações gráficas de capacitores. Apenas os termos físico-químicos são explicativos, na sua concepção.

Ainda que a biologia experimental, na qual se inclui a neurociência, possa ter autonomia quanto à organização, metodologia e procedimentos de pesquisa, ela não produz teorias próprias, apenas aplica teorias físico-químicas. Ele defende que a física descreve as leis da natureza, enquanto a biologia apenas encontra suas aplicações para os sistemas vivos. O objetivo da biologia experimental, para ele, é a redução dos fenômenos biológicos a conceitos físico-químicos. Seu ponto de vista, portanto, opõe-se ao do modelo MMM, para o qual explicar consiste em apontar as entidades e atividades de um fenômeno e como elas se organizam.

Para analisar o modelo MMM, Weber (2008) divide-o em duas partes. A primeira, que ele chama de *modelo de ação*, corresponde às passagens em que Hodgkin e Huxley elaboram a interpretação física da equação total de corrente. Weber delimita as seguintes:

- i. A despolarização causa um aumento transitório da condutância ao sódio e um mais lento, porém sustentado, aumento da condutância ao potássio (Hodgkin e Huxley, 1952, p. 500);
- ii. Essas mudanças são graduais e podem ser revertidas pela repolarização da membrana (Idem);
- iii. As mudanças da permeabilidade parecem depender do potencial de membrana, e não da corrente de membrana (Ibid., p. 501).

Hodgkin e Huxley propuseram, ainda, que a passagem dos íons pela membrana neuronal

seria feita através de um mecanismo de translocação. Esse modelo revelou-se, mais tarde, equivocado. Ainda que isso possa ser interpretado como uma evidência de que a real explicação do potencial de ação ainda estaria por vir (Bogen, 2005; Craver, 2007), para Weber, isso apenas simboliza a distinção entre as tarefas de identificar as relações causais e de revelar a base física e os mecanismos responsáveis por essas relações. As três passagens acima fariam parte do conteúdo causal do modelo e nada dizem a respeito dos mecanismos. O equívoco do mecanismo de translocação não atrapalha, de forma alguma, a validade dessa descoberta.

O *modelo de condutância*, por sua vez, consiste nas equações de condutância para o sódio e o potássio, associadas às equações que demonstram que o fluxo iônico depende da voltagem transmembrana. Ele ilustra a determinação experimental da regularidade das propriedades de condutância das membranas neurais. Weber usa a abordagem causal de Woodward (2003) para defender que esse modelo possui duas características que o tornam uma explicação causal. Em primeiro lugar, ele é invariante sob certas intervenções, como a concentração inicial do sódio e do potássio e a variação da voltagem do *clamp*. Além disso, ele apresenta modularidade, que também é uma característica das generalizações causais, segundo Woodward (2003). Isso quer dizer que intervenções em partes do modelo, como como um aumento exagerado na concentração de um íon, quebram apenas a relação às quais estão diretamente ligadas, mantendo as demais relações intactas.

Weber também argumenta contra a afirmação de Craver de que as equações do modelo HH são puramente fenomenológicas, como os epiciclos da astronomia ptolomaica, no sentido de fazerem previsões corretas mas serem, nos demais aspectos, falsos. Ele aponta que os epiciclos não são algo que poderia ser substituído por um modelo ou equação verdadeiros dentro da estrutura da astronomia pré-copernicana. Ao contrário disso, todas as equações do modelo HH podem ser reinterpretadas de maneira a fazerem sentido em conjunção com os novos conhecimentos, como dos canais iônicos e suas estruturas.

Além disso, Weber observa que, em algumas áreas da biologia, são conhecidas algumas regularidades, ainda que os mecanismos responsáveis sejam ignorados. Esse é o caso das leis de Mendel e as demais generalizações da genética clássica no início do século XX. Para Weber, o fato de eles não terem conhecimentos sobre os mecanismos genéticos desses fenômenos não quer dizer que eles não tivessem conhecimentos causais.

Por fim, Weber compara as equações do modelo HH com uma equação simples, que explica o movimento de um bloco de massa m deslizando em um plano inclinado de ângulo ϕ . Sendo g a aceleração da gravidade, μ_k o coeficiente de atrito e N a força normal exercida pelo plano, a equação pode ser dada da seguinte forma:

$$F = mg\sin\phi - \mu_k mg\cos\phi$$

Weber frisa que o próprio Woodward usa esse caso como ilustrativo de uma explicação causal. Contudo, ele aponta que esse modelo não faz qualquer menção aos mecanismos responsáveis pela força de atrito, o que envolveria conhecimentos profundos de física. O coeficiente de atrito é uma magnitude experimentalmente determinada. Embora seja possível, em princípio, calcular esse coeficiente para determinados materiais, isso de modo algum é necessário para explicar por que um bloco desliza por um plano inclinado sob determinada aceleração. Esse conhecimento apenas enriqueceria o entendimento de casos específicos. Isso também seria válido para o modelo HH. A explicação causal estaria contida nas equações e os mecanismos moleculares meramente enriqueceriam essa explicação.

Em suma, Weber defende que o modelo HH consiste em uma interligação de regularidades experimentais e leis físicas fundamentais, característica das explicações físicas. Como na física, ele se abstém de detalhes mecânicos responsáveis pelas propriedades físicas. Por fim, possuem as propriedades de invariância e modularidade, tidas como características de explicações causais. Por esse motivo, Weber conclui que ele oferece uma explicação causal para o potencial de ação, prescindindo dos mecanismos responsáveis pelo fenômeno.

O modelo HH foi elaborado com base em experimentos que avaliavam as relações entre as variações do potencial da membrana e as concentrações iônicas de sódio e potássio. Para isso, os cientistas realizaram intervenções em uma variável, medindo os efeitos sobre a outra variável. Weber tem razão: com esses experimentos, Hodgkin e Huxley estabeleceram relações que cumprem os requisitos das relações causais de Woodward, amplamente aceitas atualmente.

Weber está correto ao separar as tarefas de identificar relações causais e traçar mecanismos. No entanto, a causalidade é apenas um dos aspectos exigidos para caracterizar explicações neurocientíficas aceitáveis. A mera identificação de correlações causais não é suficiente, em neurociência, para produzir explicações. É necessário, além disso, identificar as entidades e atividades que possuem relevância causal para o fenômeno. Por esse motivo, por mais que o modelo HH descreva relações causais que permitam fazer previsões, ele não é capaz de, sozinho, explicar o potencial de ação. Isso só foi possível com os estudos subsequentes que identificaram os canais iônicos voltagem dependentes.

Por fim, digna de nota e repleta de razão é a observação de Weber sobre a comparação feita por Craver entre a astronomia ptolomaica e as equações do modelo HH. Craver diz:

O seu [de Hodgkin e Huxley] modelo das mudanças de corrente é, nesse respeito,

mais análogo aos modelos planetários ptolomaicos, que não envolvem nem implicam qualquer comprometimento com a existência dos epiciclos, deferentes e equantes a partir dos quais eles foram construídos, que ao modelo gravitacional de Newton do movimento planetário, o qual Newton apresenta para mostrar como e por que os planetas se movem como se movem (Craver, 2007, p. 54).

Weber observa que, enquanto a astronomia ptolomaica foi completamente substituída pela astronomia copernicana, as equações do modelo HH guardam seu valor integralmente, mesmo com as descobertas dos mecanismos de condução iônica. Nas suas palavras:

Primeiro, as variáveis de gatilho podem ser reinterpretadas como propriedades de canais iônicos (Kuffler, Nicholls, and Martin 1984, 149). Se às equações relevantes pode ser dada uma nova interpretação de tal modo que elas digam algo (aproximadamente) verdadeiro, então elas devem ter acertado algumas relações causais salientes, mesmo que seus autores não soubessem a sua interpretação física correta (Weber, 2008, p. 1003).

Ao defender que as equações do modelo HH são explicativas, Weber afasta-se de uma das principais tendências das abordagens explicativas em biologia: a busca por um modelo que reflita e seja coerente com a prática científica. Biólogos, incluindo neurocientistas, não procuram enquadrar fenômenos científicos em leis físicas fundamentais. Ainda que, em alguns casos particulares, como o potencial de ação, possa haver equações descritivas, preditivas, heurísticas ou mesmo prototípicas, no sentido de Schaffner, a busca por explicações só cessa quando os componentes e suas atividades são identificados.

3.4. Os mecanismos biológicos de Bechtel

A abordagem de Bechtel sobre explicações em biologia foi extensivamente discutida em Bechtel (2010). Para ele, a elaboração de explicações em biologia apresenta duas fases. A primeira delas é a decomposição dos sistemas em suas partes. Isso corresponde à descoberta de mecanismos, no mesmo sentido usado por Craver. Para Bechtel, a natureza está organizada hierarquicamente de tal forma que, em alguns sistemas, o comportamento de algumas partes componentes independe do comportamento de outras partes. Nesses casos, o comportamento das partes reflete o comportamento do sistema.

No entanto, existem alguns sistemas nos quais o comportamento das partes depende

criticamente do comportamento de outras partes. Esse tipo de sistema, a rigor, não pode ser decomposto, mas pesquisas que buscavam decompô-lo acabaram mostrando-se muito reveladoras. Um exemplo de interação entre as partes, a retroalimentação negativa⁸, é extremamente comum em biologia. Mesmo sendo um dos mecanismos de integração mais simples, houve grande dificuldade em compreendê-lo inicialmente.

Uma das características mais interessantes dos mecanismos de retroalimentação é que, se existe um atraso temporal ou interações não-lineares na retroalimentação, todo o sistema oscila. Exemplos de oscilações são as oscilações ultradianas do levedo, as oscilações na atividade cerebral e os ritmos circadianos. O comportamento de sistemas oscilantes não pode ser compreendido a partir de uma descrição de um mecanismo linear e qualitativo, como aqueles usados na abordagem MMM, em que cada mecanismo é considerado independente dos demais. Recentemente, os cientistas têm reconhecido as oscilações fisiológicas dos organismos como sinais da organização cíclica dos mecanismos. Além de serem sintomas de atividade endógena, as oscilações também são um fator importante para a manutenção dessa atividade, ao assegurarem que os organismos executem regularmente as operações necessárias para o seu reparo.

Contudo, com frequência, pesquisadores que estudam mecanismos interpretam a variabilidade dos processos endógenos como ruídos nos dados, que podem ser filtrados pela medição repetida dos fenômenos em vários momentos e vários organismos. Bechtel ressalta que, embora isso forneça dados mais homogêneos para a detecção de respostas a intervenções, também mascara os efeitos das oscilações endógenas que podem ser importantes para o próprio fenômeno em investigação. Por esse motivo, a abordagem mecanística usada por Machamer et al (2000) e Craver (2007) é insuficiente. Ainda que os autores reconheçam a possibilidade de existência de ciclos, ela é incapaz de situá-los na dinâmica endógena dos mecanismos que lhes permite ter autonomia.

Existem ainda sistemas em que os componentes estão tão uniformemente integrados que o comportamento resultante não pode ser compreendido com base no comportamento individual das partes, mas apenas quando se leva em consideração a sua organização como um todo, o que Bechtel chama de “funcionamento orquestrado do organismo”. Esse tipo de sistema tem sido cada vez mais estudado em neurociência, onde se acredita que exista um sistema com uma complexa atividade dinâmica integrada no qual os componentes individuais executam tarefas distintas.

Um das principais características que distinguem os organismos vivos dos demais é o que

⁸ Os circuitos de retroalimentação, ou *feedback*, são mecanismos de controle existentes em praticamente todos os organismos vivos, no qual um dos produtos de uma via, geralmente o produto final, entra novamente na via para controlar o seu funcionamento. A retroalimentação pode ser positiva, quando o produto estimula a atividade da via, ou negativa, quando inibe.

Bechtel chama de autonomia. Autonomia refere-se à capacidade de os organismos vivos se manterem como entidades estruturadas e duradouras diante dos processos físicos normais que tendem à desordem, o que é possível através da canalização de energia para executar certas operações de reparo necessárias. Um mecanismo no qual todos os reparos podem ser executados a partir de agentes causais situados dentro do próprio organismo requer, basicamente, o acoplamento de operações cíclicas de reparo com a energia livre decorrente de o organismo ser um sistema aberto. As operações internas precisam estar tão bem integradas umas com as outras que sejam capazes de regenerar a si mesmas.

A abordagem tradicional do modelo MMM pode ser muito bem compreendida com poucos recursos matemáticos. No entanto, para Bechtel, tão importante como decompor o sistema e localizar suas partes é recompô-lo e mostrar como a organização das partes resulta no comportamento do sistema. Em muitos casos, a recomposição literal é impossível, tornando necessário recorrer a modelamentos matemáticos ou computacionais que simulem o comportamento do todo. Em sistemas altamente integrados em que a atividade de algumas partes afeta a atividade das demais, esse recurso se torna cada vez mais necessários.

Talvez essa seja a mais bem fundamentada crítica ao modelo MMM. Ele reconhece o valor dessa abordagem, mas considera-a incompleta. Para ele, identificar as partes componentes de um fenômeno é válido, mas insuficiente, pois ignora as relações quantitativas e os ciclos de retroalimentação. Somente quando se consegue levar isso em conta e considerar os mecanismos como um todo, o que ele chama de “funcionamento orquestrado do organismo”, pode-se chegar a uma explicação completa sobre um fenômeno de um organismo vivo.

A identificação das partes componentes de mecanismos que resultam em determinados fenômenos constitui a parte mais inicial da produção de explicações em neurociência. No entanto, o entendimento completo do funcionamento cerebral só é possível quando se considera a integração desses mecanismos. Em sistemas mais simples, em que os mecanismos estão pouco integrados, a mente humana é capaz de fazer isso. No entanto, quanto mais complexo e mais interligado é o sistema, mais necessário se torna recorrer a modelamentos para gerar uma explicação satisfatória.

3.5. Mecanismos e relevância constitutiva

A última crítica ao modelo MMM que examinaremos é a de Couch (2011). Embora não critique as bases estruturais do modelo MMM, ele chama atenção para a falta de clareza de alguns pontos. Para ele, “a noção de mecanismo ainda permanece inadequadamente compreendida” (Couch, 2008, p. 375), o que se refere na falta de critérios para distinguir partes relevantes e

irrelevantes de um mecanismo.

Para ilustrar seu ponto de vista, Couch usa o mecanismo de discriminação de cores pelo olho humano. É interessante que, num artigo em que ele se propõe a falar da importância de distinguir partes relevantes e irrelevantes, Couch comete alguns equívocos e omite partes fundamentais para o mecanismo, citando outras partes irrelevantes. Ele diz: “a informação sobre a luz do ambiente passa pela pupila e impacta receptores da retina no olho posterior (bastonetes e cones)” (Couch, 2008, p. 376). Os bastonetes, embora sejam extremamente importantes para a visão em ambientes com pouca luminosidade, não têm qualquer papel na discriminação de cor. Ele comete outro equívoco quando diz: “os receptores da retina contêm diferentes tipos de moléculas de rodopsina”. A rodopsina é, na realidade, a proteína fotossensível dos *bastonetes*. Há três diferentes tipos de cones, cada um deles com um tipo de pigmento sensível a uma determinada faixa de comprimento de onda: curta (pigmento B), média (pigmento G) e longa (pigmento R). Por fim, ele deixa de citar o mais importante na percepção da cor: os cones individuais não transmitem informações acerca do comprimento de onda de um estímulo luminoso. A percepção da cor é feita através da comparação das respostas dos três tipos de cones, no cérebro (Kandel, 1997).

Couch defende que o critério de Craver, baseado na abordagem manipulacionista da causalidade, é meramente epistêmico. Nesse sentido, ele apenas oferece uma forma de verificar se uma parte é relevante ou não para o mecanismo, nada dizendo sobre a relação de relevância, uma relação que se dá entre as partes do mecanismo e a atividade que ele executa. Nesse sentido, Couch defende que esse tipo de abordagem causal confunde problemas epistemológicos com a própria natureza das relações causais.

Couch propõe uma abordagem que permita diferenciar partes necessárias das desnecessárias, olhando para a própria natureza dessa relação. Essa abordagem é semelhante à noção de condição *inus* de Mackie (1974), citada em Couch (2011): considere-se que um evento qualquer tenha sido causado por uma condição complexa, composta por múltiplas condições individuais. Nesse caso, cada uma das condições é uma *parte insuficiente, mas não redundante, de uma condição desnecessária, mas suficiente (insufficient but nonredundant part of an unnecessary but sufficient condition – inus condition)*. Por exemplo, em uma casa ocorre um curto-circuito e, em decorrência dele, a casa pega fogo. O curto-circuito é, nesse caso, a causa para o incêndio da casa, assim como as outras condições presentes, como o oxigênio e o material combustível.

Couch acredita que uma analogia pode ser aplicada para a determinação da relevância causal em um mecanismo. Os componentes do mecanismo seriam, então, partes de uma estrutura complexa que servem para dar conta do fenômeno. Cada componente, individualmente, é necessário para a presença da capacidade, mas insuficiente para a sua ocorrência. Nesse sentido,

cada parte relevante do mecanismo pode ser vista como um componente *inus*.

Parece, contudo, que a abordagem de Mackie não resolve o problema que Couch quis solucionar. Isso decorre de uma confusão na própria noção de mecanismo. Couch diz: “*o olho humano é um mecanismo que contém muitas partes que contribuem para outras capacidades além da discriminação de cores, como a discriminação de brilho, percepção da distância, detecção do movimento, etc. É compreendido que nem toda parte ou componente do olho é envolvido em cada uma das capacidades executadas*” (Couch, 2011, p. 377, grifo meu).

No entanto, o olho humano *não é*, segundo a abordagem mecanística, um mecanismo. Mecanismo diz respeito a um processo, à própria execução da capacidade. Não há um mecanismo “olho”, mas um mecanismo da percepção da cor, que seria o seguinte: “Sabe-se que a luz passa pela pupila e chega aos receptores situados na retina, os cones e bastonetes. Existem três tipos de cones, cada um deles com um tipo de pigmento sensível a uma determinada faixa de comprimentos de onda: curta (pigmento B), média (pigmento G) e longa (pigmento R). Quando é ativado, o cone transduz o sinal luminoso em sinal elétrico, que é projetado para as células ganglionares através das células bipolares. Os axônios das células ganglionares formam o nervo óptico, que se projeta para o cérebro. O cérebro, então, compara as respostas dos três tipos de cones e, com isso, detecta a cor”.

Nesse caso, a noção de condição *inus* proposta por Couch não resolve os problemas a que se propõe. Ela permite considerar relevantes partes que são, de fato, irrelevantes, como a bainha de mielina⁹ e o citoplasma neuronal, que seriam análogos ao oxigênio e o material combustível do exemplo de Couch. Afinal, se não houvesse bainha de mielina ou se o neurônio estivesse repleto de outro líquido diferente do citoplasma, não haveria percepção de cor.

Craver (2007) parece ter razão. Segundo ele, “a busca por essa conexão [que Hume buscava entre a causa e o efeito] (...) é mal orientada e frequentemente distrai filósofos dos aspectos da causação que são mais importantes para uma abordagem da explicação. Desde o início, ele ressalta a conexão íntima que existe entre relevância causal, explicação e controle. Sua abordagem é epistemológica porque, na sua concepção, não é possível separar esses aspectos da explicação em neurociência.

9 A bainha de mielina consiste em camadas concêntricas de lipídios alternados entre camadas protéicas envolvendo os neurônios e que serve para aumentar a velocidade de condução do impulso nervoso.

Conclusão

A neurociência é, hoje, uma área autônoma e unificada da biologia experimental. Por esse motivo, ela formula problemas, possui pressupostos, conduz experimentos e elabora explicações de um modo particular. As discussões sobre explicações em neurociência são muito recentes e, frequentemente, feitas no contexto de uma discussão mais geral. Explicações nesse domínio eram apenas usadas como exemplos para sustentar modelos elaborados para a biologia de forma geral ou biologia experimental. Por esse motivo, as questões particulares da neurociência eram deixadas de lado em favor de uma discussão mais ampla.

Apenas nos últimos vinte anos, mais ou menos, as discussões sobre explicações em neurociência adquiriram certa independência. Em grande parte, isso se deveu aos trabalhos de Carl Craver publicados entre 2000 e a atualidade. Eles deram origem ao modelo explicativo mais expressivo que existe atualmente para a neurociência: o modelo mecanístico-causal multinível (MMM).

A formulação desse modelo pretendeu satisfazer duas necessidades em filosofia da neurociência: em primeiro lugar, descrever a forma como neurocientistas elaboram suas explicações. Segundo, oferecer critérios para avaliar se novas explicações formuladas podem ser consideradas boas explicações. Desse modo, podem ser distinguidos dois objetivos do modelo MMM: um objetivo normativo e um objetivo descritivo, havendo sempre uma relação estreita entre esses dois objetivos.

Para tanto, foram examinados alguns casos incontroversos de explicações neurocientíficas. Particularmente, foram estudados o mecanismo do potencial de ação, o mecanismo da sensitização de longo prazo e o mecanismo de formação da memória. Diante de um exame detalhado, procurou-se verificar o que essas explicações têm em comum e que as distinguem de outras explicações em ciência geral e em biologia.

O primeiro aspecto em que a neurociência se destaca é que ela procura construir explicações mecanísticas. O modelo nomológico-dedutivo, como vimos, é insatisfatório para caracterizar as explicações neurocientíficas. O principal motivo é que neurocientistas não concebem suas explicações como argumentos. De modo geral, eles revelam um comprometimento com a existência ôntica dos fenômenos que eles descrevem, assim como dos mecanismos e as entidades e atividades que o compõem. O modelo nomológico dedutivo também não é capaz de lidar com certas questões que neurocientistas buscam desvendar, especificamente a relevância causal. Uma vez que a explicação é construída na forma de argumento, é impossível determinar quais as partes do *explanans* são relevantes para a ocorrência do *explanandum*. Por fim, o modelo nomológico-

dedutivo ignora a diferença entre explicação e previsão. Uma vez que muitos fenômenos neurocientíficos são estocásticos e imprevisíveis, explicar e prever são atividades bastante diferentes.

O sentido de mecanismo que é usado aqui é distinto da noção clássica de mecanismo, que remete ao funcionamento de uma máquina. Mecanismo refere-se a entidades e atividades organizadas de tal forma que produzem o fenômeno. Uma explicação mecanística adequada precisa conter todas as partes que são causalmente relevante para o mecanismo, e nada além delas. Quando as partes não estão todas descritas, tem-se apenas um esboço do mecanismo. A atividade do neurocientista é completar as lacunas desse esboço até a descrição do mecanismo completo.

Assim, explicações neurocientíficas são descrições de mecanismos que têm como objetivo situar um fenômeno na estrutura causal do mundo. Elas são, portanto, explicações causais, com comprometimento com a existência ôntica dos fenômenos. Elas necessitam de uma abordagem de causalidade adequada, que seja capaz de lidar com as peculiaridades da neurociência, levando em conta que os fenômenos neurocientíficos são limitados em abrangência, estocásticos, mecanisticamente frágeis e historicamente contingentes. Além disso, é necessário considerar a existência de causação negativa, por omissão e prevenção, que estão presentes em diversas explicações elaboradas por neurocientistas. Por fim, uma abordagem causal em neurociência precisa levar em conta o aspecto pragmático dessa área, ou seja, de que intervir no cérebro para tratar doenças e melhorar o seu funcionamento são indissociáveis da prática neurocientífica.

A maioria das abordagens causais que existem atualmente foi elaborada para lidar com questões referentes à física. Por esse motivo, elas ignoram aspectos particulares de algumas áreas da ciência, incluindo a neurociência, cujos fenômenos são limitados em abrangência, estocásticos, mecanisticamente frágeis e historicamente contingentes. Mais recentemente, a abordagem manipulacionista da causalidade foi formulada especificamente para lidar com essas áreas, particularmente as ciências sociais e biomédicas. Por esse motivo, essa formulação é a mais adequada para constituir explicações em neurociência.

De acordo com a abordagem manipulacionista da causalidade, relações causais são distintas de outras relações de contiguidade na medida em que são potencialmente exploráveis para o propósito de manipulação e controle. Explicações causais descrevem relações entre variáveis que podem ser exploradas para produzir, prevenir e alterar o *explanandum*. Elas exibem um padrão de dependência contrafactual entre o *explanans* e o *explanandum*, mostrando como intervenções nas variáveis do *explanans* resultam em alterações do *explanandum*. Além disso, a relação entre *explanans* e *explanandum* deve ser invariante, ou seja, deve haver uma generalização capaz de descrever como intervenções no *explanans* alteram o *explanandum*.

Outro aspecto importante das explicações mecanísticas é que elas possuem componentes causalmente relevantes pertencentes a vários níveis. Uma abordagem multinível vai de encontro com teóricos como Bickle, que acreditam que o avanço das pesquisas em níveis fundamentais tornará obsoletos os níveis mais altos. Essa concepção de nível diz respeito a níveis de composição, em que componentes de um nível mais baixo estão organizados para formar componentes de níveis mais altos, relacionando-se uns com os outros para produzir o fenômeno a ser descrito pelo mecanismo. Ainda que componentes de nível mais alto possuam *relevância* causal para o fenômeno, ou seja, fazem diferença na ocorrência do *explanandum*, isso é distinto de *poder* causal, num sentido metafísico. Por esse motivo, o argumento da completude causal no nível fundamental não se aplica a essa abordagem.

A última particularidade da neurociência é definir como ela se estrutura como uma ciência unificada. Como vimos, modelos clássicos de redução interteórica não se mostraram adequados, já que tipos-termos das diferentes áreas não podem ser traduzidos para a linguagem de outras. Ao invés disso, o modelo MMM propõe que a unidade da neurociência é obtida quando os cientistas das várias áreas que a compõem trabalham de forma complementar para construir mecanismos.

A discussão sobre mecanismos e multinivellaridade gerou uma vasta discussão entre filósofos da ciência. Diferente de modelos anteriores, o modelo MMM não procura normatizar a atividade do cientista com base em normas estabelecidas para uma outra prática científica. Ao contrário, esse modelo se fundamenta na prática dos neurocientistas e nos resultados e explicações conforme apresentados por eles. Por esse motivo, ele constitui a abordagem mais coerente com a neurociência na atualidade.

No entanto, ele não está livre de falhas, ainda que estas não comprometam suas bases. Um dos aspectos é que o modelo MMM precisa lidar com o fato de que os mecanismos qualitativos a que ele se refere são *protótipos* de explicação, que captam as generalidades de um fenômeno, mas que podem ocorrer com ligeiras modificações em diferentes seres vivos.

Além disso, o modelo MMM possui uma limitação importante. O cérebro possui uma estrutura integrada, em que mecanismos não ocorrem de forma independente e ordenada, mas interferem uns nos outros a todo momento. Esse funcionamento coordenado dos mecanismos gera oscilações temporais que são importantes para a manutenção da autonomia do organismo. A mera descrição de mecanismos não é capaz de captar esse fenômeno integral do funcionamento cerebral, limitando-se a fornecer descrições separadas de diversos mecanismos. Os modelamentos computacionais e matemáticos permitem reproduzir esse funcionamento orquestrado e podem ser uma possibilidade de compreender a integração dos mecanismos cerebrais.

Ainda assim, o modelo MMM constitui a abordagem mais completa, detalhada e coerente

com a prática neurocientífica que existe até o momento. Acreditamos que, se suas críticas forem consideradas seriamente, ele pode se tornar ainda melhor, lidando com suas limitações e ganhando ainda mais robustez.

Referências bibliográficas

ABRANTES, P. C. *Filosofia da Biologia*. Porto Alegre: Artmed, 2011. 326p.

ARMSTRONG, C. M. Sodium Channels and Gating Currents. *Physiological Reviews*, 61(3), p. 644-683, 1981.

BARNES, E. Explanatory Unification and the Problem of Assymetry. *Philosophy of Science*, 59 p. 558-571, 1992.

BECHTEL, W., ABRAHAMSEN, A. Explanation: A Mechanistic Alternative. *Studies in the History and Philosophy of the Biological and Biomedical Sciences*, 36, p. 421-441, 2005.

BECHTEL, W., RICHARDSON, R. C. *Discovering Complexity: decomposition and localization as strategies in scientific research*. Cambridge: Princeton University Press, 2010, 286p.

BICKLE, J. *Philosophy and Neuroscience*. Holanda, Kluwer Academic Publishers, 2003, 235p.

BICKLE, J. Reducing Mind to Molecular Pathways: explicating the reductionism implicit in current cellular and molecular neuroscience. *Synthese*. 151:411-434, 2006.

BLISS, T. V. P., GARDNER-MEDWIN, A. R., LOMO, T. Synaptic Plasticity in the Hippocampal Formation. In: ANSELL, G. B., BRADLEY, P. B. *Macromolecules and Behavior*. London: Macmillan, 1973. p. 193-203.

BLISS, T. V. P., GARDNER-MEDWIN, A. R. Long Lasting Potentiation of Synaptic Transmission in the Dentate Area of Unnaesthetized Rabbit Following Stimulation of the Perforant Path. *Journal Physiology*. 232: 357-374, 1973.

BLISS, T. V. P., COLLINGRIDGE, G. L. A Synaptic Model of Memory: Long Term Potentiation in the Hippocampus. *Nature*. 361/6407: 31-39, 1993.

BOGEN, J. The Hodgkin and Huxley Equations and the Concrete Model: Comments on Craver, Schaffner, and Weber. *Philosophy of Science*, 75(5), 1034-1046, 2008.

CHURCHLAND, P. M. *A Neurocomputational Perspective: the nature of mind and the structure of science*. Cambridge: MIT Press, 341 p.

COFFA, J. A. Probabilities: Reasonable or True. *Philosophy of Science*. 44 p. 186-198, 1977.

COUCH, M. B. Mechanisms and Constitutive Relevance. *Synthese*.183, p. 375-388, 2011.

CRAVER, C. *Explaining the Brain: mechanisms and the mosaic unity of neuroscience*. New York: Oxford University Press Inc., 2007. 308 p.

CUMMINS, R. Functional Analysis. *The Journal of Philosophy*, 72, p. 741-765, 1975.

DALE, H.H. et al. Release of Acetylcholine at Voluntary Motor Nerve Endings. *Journal of Physiology*, 86, p. 353-380, 1936.

DOWE, P. Wesley Salmon's Process Theory of Causality and the Conserved Quantity Theory. *Philosophy of Science*, 59, p. 195-216, 1992.

DOWE, P. *Physical Causation*. Cambridge: Cambridge University Press, 2000. 248 p.

DOWE, P. Causes are Physically Connected to their Effects: Why Preventers and Omissions are Not Causes. In: HITCHCOCK, C. (ed). *Contemporary Debates in Philosophy of Science*. Oxford: Blackwell, 2004. p. 189-196.

DRETSKE, F. If You Can't Make One, You Don't Know How It Works. In: FRENCH, P., UEHLING, JR., WETTSTEIN, H. (eds). *Midwest Studies in the Philosophy, 19, Philosophical Naturalism*. Notre Dame: University of Notre Dame Press, 1994.

GLENNAN, S. S. Mechanisms and the Nature of Causation. *Erkenntnis*, 44, p. 49-71, 1996.

GOODMAN, N. The Infirmities of Confirmation Theory. *Philosophy and Phenomenological Research*, 8 p. 149-151, 1947.

HEMPEL, C. G., OPPENHEIM, P. (1948). Studies in the Logic of Explanation. *Philosophy of Science*, 15, p. 135-175, 1965.

HEMPEL, C. G., OPPENHEIM, P. Deductive Nomological vs. Statistical Explanation. In: FEIGL, H., MAXWELL, G. *Minnesota Studies in the Philosophy of Science III*. Minneapolis: University of Minnesota Press, 1962. p. 98-169.

HEMPEL, C. G. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: The Free Press. 1965. 505 p.

HILLE, B. *Ion Channels of Excitable Membranes*. Sunderland: Sinauer Associates, 1984. 814 p.

HILLE, B., ARMSTRONG, D., MACKINNON, R. Ion Channels: From Idea to Reality. *Nature Medicine*, 5, p. 1115-1119, 1999.

HITCHCOCK, C. Discussion: Salmon on Explanatory Relevance. *Philosophy of Science*, 62 p. 304-320, 1995.

HODGKIN, A. L., HUXLEY, A. F. Action Potentials Recorded from Inside a Nerve Fibre. *Nature*, 144, 710-711, 1939.

HODGKIN, A. L., HUXLEY, A. F. Measurement of Current-Voltage Relations in the Membrane of the Giant Axon of Loligo. *Journal of Physiology*, 116, 424-448, 1952a.

HODGKIN, A. L., HUXLEY, A. F. Currents Carried by Sodium and Potassium Ions Through the Membrane of the Giant Axon of Loligo. *Journal of Physiology*, 116, 449-472, 1952b.

HODGKIN, A. L., HUXLEY, A. F. The Components of Membrane Conductance in the Giant Axon of Loligo. *Journal of Physiology*, 116, 476-496, 1952c.

HODGKIN, A. L., HUXLEY, A. F. The Dual Effect of Membrane Potential on Sodium Conductance in the Giant Axon of Loligo. *Journal of Physiology*, 116, 497-506, 1952d.

HODGKIN, A. L., HUXLEY, A. F. A Quantitative Description of Membrane Current and its Application to Conduction and Excitation in Nerve. *Journal of Physiology*, 117, 500-544, 1952e.

KANDEL E. R., SCHWARTZ, J. H., JESSEL, T. M. , S. (1995). **Fundamentos da Neurociência e do Comportamento**. Tradução por Charles Alfred Esbérard e Mira de Casrilevitz Engelhardt. 2. ed. Rio de Janeiro: Guanabara Koogan, 2007.

KATSUKI T. et al. Compartmentalization within neurites: its mechanisms and implications. *Developmental Neurobiology*. 71(6):458-73, 2011.

KATZ, B., MILEDI, R. The Release of Acetylcholine from Nerve Endings by Graded Electrical Pulses. *Proceedings of the Royal Society*, 167, p. 23-38, 1967.

KATZ, B., MILEDI, R. The Timing of Calcium Action during Neuromuscular Transmission. *Journal of Physiology*, 189, 535-544, 1967.

KIM, J. Mind in a Physical World. Cambridge: MIT Press, 1998.

KITCHER, P. 1953 and All That. A Tale of Two Sciences. *The Philosophical Review*, 93, p. 335-373, 1984.

KITCHER, P. Explanatory Unification and the Causal Structure of the World. In: KITCHER, P., SALMON, W. C. *Scientific Explanation*. Minneapolis: University of Minnesota Press, 1989. p. 410-505.

LUPO G. et al. Retinoic acid receptor signaling regulates choroid fissure closure through independent mechanisms in the ventral optic cup and periorcular mesenchyme. *Proceedings of the National Academy of Sciences*. Maio 2011, no prelo.

MACHAMER, P. K., DARDEN, L., CRAVER, C. F. Thinking About Mechanisms. *Philosophy of Science*, 57, 1-25, 2000.

MILEDI, R. Transmitter Release Induced by Injection of Calcium Ions into Nerve Terminals. *Proceedings of the Royal Society of London*, 183, 421-425, 1973.

MUNSON, R. Why Medicine Cannot be a Science. *The Journal of Medicine and Philosophy*, 6 p. 183-208, 1981.

NAGEL, E. The Meaning of Reduction in the Natural Sciences. In: STAUFFER, R. C. *Science and Civilization*. Madison: University of Wisconsin Press, 1949. p. 97-135.

NAGEL, E. *The Structure of Science: Problems in the Logic of Scientific Explanation*. New York: Harcourt, Brace & World, 1971. 618 p.

OBERLANDER J. G. Anabolic androgenic steroid abuse: multiple mechanisms of regulation of gabaergic synapses in neuroendocrine control regions of the rodent forebrain. *Journal of Neuroendocrinology*. Maio 2011, no prelo.

OPPENHEIM, P., PUTNAM, H. Unity of Science as a Working Hypothesis. In: FEIGL, H., SCRIVEN, M., MAXWELL, G. *Minnesota Studies in the Philosophy of Science II*. Minneapolis: University of Minnesota Press, 1958. p. 3-36.

RESCHER, N. The Stochastic Revolution and the Nature of Scientific Explanation. *Synthese*, 14, p. 200-215, 1962.

SALMON, W. C. *Statistical Explanation and Statistical Relevance*. Pittsburgh: University of Pittsburgh Press, 1971. 128 p.

SALMON, W. C. An At-At Theory of Causal Influence. *Philosophy of Science*, 44 p. 215-224, 1977.

SALMON, W. C. Why Ask Why? - An Inquiry Concerning Scientific Explanation. *Proceedings and Addresses of the American Philosophical Association*, 51 p. 683-705, 1978.

SALMON, W. C. *Scientific Explanation and the Causal Structure of the World*. Princeton:

Princeton University Press, 1984. 321 p.

SALMON, W. C. Scientific Explanation. In: SALMON, M. *Introduction to the Philosophy of Science*. Englewood Cliffs: Prentice Hall, 1992. p. 7-41.

SALMON, W. C. Causality Without Counterfactuals. *Philosophy of Science*, 61 p. 297-312, 1994.

SALMON, W. C. Causality and Explanation: A Reply to Two Critiques. *Philosophy of Science*, 64 p. 461-477, 1997.

SALMON, W. C. *Causality and Explanation*. New York: Oxford University Press, 1998. 448 p.

SALMON, W. C. *Four Decades of Scientific Explanation*. Pittsburgh: First University of Pittsburgh Press, 2006. 234 p.

SCOVILLE, W. B., MILNER, B. Loss of Recent Memory After Bilateral Hippocampal Lesions. *Journal of Neurology, Neurosurgery and Psychiatry*, 20, p. 11-21, 1957.

SCHAFFER, J. Metaphysics of Causation. In: STANFORD Encyclopedia of Philosophy. Stanford: The Metaphysics Research Lab, 2003.

SCHAFFNER, K. F. *Discovery and Explanation in Biology and Medicine*. Chicago: The University of Chicago Press, 1993, 617p.

SCHAFFNER, K. F. Philosophy of Medicine. In: SALMON, M. *Introduction to the Philosophy of Science*. Englewood Cliffs: Prentice Hall, 1992. p. 310-345.

SCHAFFNER, K. F. Theories, Models and Equations in Biology: the Heuristic Search for Emergent Simplifications in Neurobiology. *Philosophy of Science*, 75, p. 1008-1021, 2008.

SCRIVEN, M. Explanation and Prediction in Evolutionary Theory. *Science*, 30, p. 477-482, 1959.

SOBER, E. *Evidence and Evolution: The Logic Behind the Science*. Nova Iorque: Cambridge University Press, 2008.

SOCIETY OF NEUROSCIENCE. *What is Neuroscience?* Disponível em : <http://www.sfn.org/index.aspx?pagename=whatIsNeuroscience>. 30/03/2012

TOLMAN, E. C. There is More Than One Kind of Learning. *The Psychological Review*, 56, p. 144-155, 1949.

WEBER, M. *Philosophy of Experimental Biology*. Nova Iorque: Cambridge University Press, 2005. 358p.

WEBER, M. Causes Without Mechanisms: experimental regularities, physical laws and neuroscientific explanation. *Philosophy of Science*, 75, p995-1007, 2008

WIMSATT, W. C. Reductive Explanation: A Functional Account. In: COHEN, R. S. ET AL *PSA 1974*. Dordrecht: Reidel, 1976. p. 671-710.

WIMSATT, W. C. The Ontology of Complex Systems: Levels, Perspectives, and Causal Thickets. *The Canadian Journal of Philosophy (Sup)*, 20, p. 207-274, 1994.

WOODWARD, J. *Making Things Happen: a theory of causal explanation*. New York: Oxford University Press Inc., 2003. 410 p.

WOODWARD, J., Scientific Explanation. In: STANFORD Encyclopedia of Philosophy. Stanford: The Metaphysics Research Lab, 2010.