

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE FILOSOFIA E CIÊNCIAS HUMANAS
CURSO DE FILOSOFIA**

Augusto Lucas Valmini

**INTERVENCIONISMO CAUSAL: EXPLICAÇÕES E
RELAÇÕES CAUSAIS**

PORTO ALEGRE

2015

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE FILOSOFIA E CIÊNCIAS HUMANAS
CURSO DE FILOSOFIA**

Augusto Lucas Valmini

**INTERVENCIONISMO CAUSAL: EXPLICAÇÕES E
RELAÇÕES CAUSAIS**

Trabalho de Conclusão de Curso, apresentado à banca examinadora da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do grau de bacharel em Filosofia, sob a orientação do Prof. Dr. Eros Moreira de Carvalho.

PORTO ALEGRE

2015

SUMARIO

INTRODUÇÃO	5
1 O MODELO NOMOLÓGICO-DEDUTIVO	6
1.1 APRESENTAÇÃO E PROBLEMAS DO MODELO NOMOLÓGICO-DEDUTIVO....	6
2 A TEORIA INTERVENCIONISTA DA CAUSALIDADE	11
2.1 LOCALIZAÇÃO DA TEORIA INTERVENCIONISTA.....	11
2.2 MOTIVAÇÃO PARA UMA ABORDAGEM INTERVENCIONISTA.....	15
2.3 APRESENTAÇÃO INFORMAL DA TEORIA.....	19
2.4 TEORIA INTERVENCIONISTA.....	24
2.5 INTERVENÇÃO CARACTERIZADA PRECISAMENTE.....	31
2.6 REDUACIONISMO.....	32
2.7 INVARIÂNCIA E LEIS.....	34
2.8 ALTERANDO O VALOR DE UMA VARIÁVEL.....	38
3 ENFRENTANDO AS CRÍTICAS DO MODELO NOMOLÓGICO-DEDUTIVO	42
3.1 INTRODUÇÃO.....	42
3.2 CRÍTICAS FEITAS AO MODELO NOMOLÓGICO-DEDUTIVO.....	43
3.2.1 Assimetria Explicativa.....	43
3.2.2 Irrelevância Explanatória.....	43
3.2.3 Alegações Causais Singulares.....	44
3.2.4 Leis e Ciências Especiais.....	44
4 CONCLUSÃO	46
REFERÊNCIAS BIBLIOGRÁFICAS	48
ANEXO	50

RESUMO

O objetivo desta monografia é argumentar em favor da utilização de uma teoria causal para fornecer uma maneira de compreendermos melhor explicações. Para nossa argumentação, expomos o modelo nomológico-dedutivo de explicação que tradicionalmente centralizou o debate sobre modelos explicativos e também os contraexemplos ao modelo. Após isso, apresentamos uma teoria intervencionista da causação que propõem os critérios que uma relação deve possuir para ser causal. Para a abordagem intervencionista da causação, uma explicação é melhor do que outra quando responde a um número maior de questões sobre o que teria acontecido se as coisas fossem diferentes. Por fim, nós usamos a teoria intervencionista para resolver os contraexemplos ao modelo nomológico-dedutivo, de modo que possamos incorporar os benefícios do modelo.

Palavras chave: Causação, Modelos Explicativos, Epistemologia, Metafísica.

ABSTRACT

The present study aims to argue in favor of a causal theory capable of providing a way to better understand explanations. For our discussion, we present the nomological-deductive model of explanation, which traditionally attracted the attention in discussions about explanation models, and also the counterexamples to this model. After that, we present an interventionist theory of causation that proposes which are the conditions a relation must meet to be causal. To an interventionist account of causation, an explanation is better than another when it can answer a larger number of questions about what would happen if things were different. Last, we use the interventionist account of causation to show how we can answer the nomological-deductive counterexamples, in a way that it incorporates the benefits of the model.

Keywords: Causation, Explanantion Models, Epistemology, Metaphysics.

INTRODUÇÃO

Nosso objetivo nesta monografia é argumentar em favor da utilização de uma teoria causal intervencionista para fornecer uma maneira de compreendermos melhor explicações causais; não nos ocuparemos, portanto, com explicações apenas descritivas, como “sal é branco”, ou explicações fornecidas por ciências formais, como “P ou Q, não P, logo Q”.

O primeiro capítulo consiste de uma exposição do modelo nomológico-dedutivo e das principais críticas que lhe foram feitas, as quais nos ocuparemos de responder no terceiro capítulo. Escolhemos expor tal modelo por ter centralizado e influenciado grande parte das discussões subsequentes sobre modelos explicativos.

O segundo capítulo apresenta a teoria causal intervencionista proposta por Woodward (2003). Nosso objetivo com ela é darmos um tratamento à noção de causa, o que permitirá criar critérios qualificatórios para explicações.

O terceiro capítulo consiste na aplicação do arcabouço teórico desenvolvido no segundo capítulo com fins de enfrentar as críticas levantadas contra o modelo nomológico-dedutivo.

Na conclusão destacamos alguns pontos interessantes que decorrem do percurso argumentativo feito aqui.

Deixo as seguintes sugestões para leitores com diferentes interesses: i) leitores mais interessados no papel que a causalidade ocuparia em explicações, e menos interessados na teoria causal, podem se desinteressar pela leitura da seção 2.4 que apresenta as caracterizações formais da teoria, e neste caso adiar sua leitura. As caracterizações são importantes para a expressividade da teoria causal e permitem lidar com contraexemplos que surgiriam de uma caracterização informal, mas tais caracterizações não são extensivamente usadas no terceiro capítulo; ii) já leitores mais interessados no que seria uma teoria causal intervencionista, e menos interessados nas questões epistemológicas que podem ser por ela tratadas, podem ler primeiramente o segundo capítulo e adiar a leitura dos outros.

Usamos “causalidade” e “causalidade” como sinônimos.

CAPÍTULO 1 – O MODELO NOMOLÓGICO-DEDUTIVO

1.1 Apresentação e Problemas do Modelo Nomológico-Dedutivo

O modelo nomológico-dedutivo (ND, para abreviar) proposto por Hempel (1974, pp. 68-73) defende que uma explicação científica deve consistir de 1a) um conjunto de leis e 1b) um conjunto de condições iniciais, que conjuntamente servem de premissas a 2) uma conclusão que segue-se dedutivamente das premissas. 1a) e 1b) são o *explanans* da explicação – a explicação do fenômeno; enquanto 2) é o *explanandum* – o fenômeno a ser explicado. As premissas do *explanans* devem ser verdadeiras e essenciais ao *explanandum*: se uma premissa fosse removida o *explanandum* não mais seria deduzido do *explanans*. Um dos principais atrativos do modelo ND é que ao variarmos as condições no *explanans* podemos prever o que ocorrerá ao *explanandum* devido ao caráter nômico de ao menos uma das premissas; existe então expectabilidade nômica do *explanandum*. Com essa possibilidade de predição capturamos algo que é muito importante, e talvez essencial, em modelos explicativos: explicamos o porquê de algo ter ocorrido, em oposição a meramente descrever o que ocorreu. Explicações científicas, mais do que explicações da vida ordinária, teriam esse caráter distintivo de previsibilidade que é capturado de maneira explícita pelo modelo ND. O exemplo fornecido por Hempel é o seguinte:

- a) Em qualquer local, a pressão exercida na sua base pela coluna de mercúrio no tubo de Torricelli é igual à pressão exercida na superfície livre do mercúrio existente na cuba pela coluna de ar acima dela.
- b) As pressões exercidas pelas colunas de mercúrio e de ar são proporcionais aos seus pesos; e quanto menor a coluna menor o seu peso.
- c) A coluna de ar acima da cuba aberta é certamente menor quando o aparelho está no alto da montanha do que quando está em baixo
- d) (Portanto), a coluna de mercúrio no tubo é certamente menor quando o aparelho está no alto da montanha do que quando está em baixo. (idem, p. 68)

Neste exemplo a), b) e c) são o *explanans* e d) o *explanandum*; a) e b) são leis gerais e c) descreve um fato particular; enquanto d) segue-se dedutivamente de a), b) e c). Se variássemos a condição inicial c) alteraríamos o *explanandum* d).

Hempel também forneceu dois modelos de explicações que não se adequariam diretamente ao modelo ND. Explicações estatístico-dedutivas (idem, pp. 78-88) que envolveriam a dedução da probabilidade de um fenômeno dadas leis probabilísticas, por exemplo das leis da teoria física corrente poderíamos deduzir que “a vida média do átomo

polônio²¹⁸ é de 3,05 minutos” (idem., p. 87). E haveriam também explicações estatístico-indutivas (idem, pp. 89-95) que tentariam deduzir de leis probabilísticas a probabilidade de um fenômeno particular, por exemplo, qual seria a probabilidade de um determinado átomo de polônio²¹⁸ se desintegrar no próximo minuto. Falaremos apenas das explicações ND por dois motivos. Primeiro, porque essas duas versões do modelo ND são variações dele, e são explicativas tomando o que há de essencial do modelo ND (a derivação de um *explanandum* de determinado *explanans*). E segundo, porque a exposição da teoria causal que faremos não se ocupará de explicações probabilísticas ou de eventos particulares.

Dividimos aqui as críticas ao modelo ND entre internas e externas. As críticas externas recusam a adequação do modelo às ciências especiais (como Economia, Biologia e Psicologia) porque as condições não se fazem presentes em explicações consideradas bem adequadas dentro das ciências especiais. Já as críticas internas são os contraexemplos às condições de necessidade e suficiência dos critérios de uma explicação. Contraexemplos à suficiência vão em duas direções: assimetria explicativa e irrelevância explicatória.

A assimetria explicativa ocorre em casos em que das leis (L) e condições iniciais (I) do *explanans* chegamos à um *explanandum* (E) tal que, se no *explanans* substituíssemos I por E, chegaríamos a um *explanandum* I. Exemplificando, podemos usar o tamanho de um mastro, a inclinação do sol relativa ao mastro e a lei a sobre propagação de luz para derivarmos o tamanho da sombra do mastro. Inversamente, poderíamos utilizar o tamanho da sombra para calcular o tamanho do mastro. Ambas explicações atendem aos critérios do modelo ND, mas enquanto diríamos que o tamanho do mastro explica porque a sombra tem determinado tamanho, não diríamos que o tamanho da sombra explica porque o mastro tem determinado tamanho. A intuição para isso é que a sombra é causada pela mastro, e não o oposto. A irrelevância explicatória ocorre quando as leis (L) conjuntamente com as condições iniciais (I) parecem irrelevantes ao *explanandum* (E). Em um exemplo de Salmon (1971, p. 34):

L: Todos homens que tomam pílulas anticoncepcionais não conseguem engravidar.

I: Jones é um homem que toma pílulas anticoncepcionais.

E: Jones não consegue engravidar.

O argumento acima é dedutivamente válido e inclui uma lei, mas apesar do *explanandum* seguir-se do *explanans*, não parece que aquele é explicado por esse. Novamente, o problema parece envolver causalidade: o *explanans* não expressa a causa de Jones não conseguir engravidar.

Já os contraexemplos à necessidade dos critérios do modelo ND recorrem a casos onde alegações singulares parecem explicativas sem recorrer a leis ou argumentos dedutivamente

válidos. Em exemplo fornecido por Scriven (1962, p. 456 apud WOODWARD 2003, p. 156):

(1.1) O impacto do meu joelho com a mesa causou a queda do tinteiro.¹

Scriven argumenta que uma alegação como (1.1) é de transparência primitiva; um homem das cavernas poderia entender seu conteúdo e usá-lo para explicações (caso perguntassem porque o tapete sob a mesa manchou) mesmo que desconhecesse os critérios do modelo ND. Uma resposta à crítica de Scriven, inclusive a de Hempel (1965), é que tomamos (1.1) como verdadeira sem uma explicação que satisfaça os critérios do modelo ND porque (1.1) nos é uma alegação causal muito familiar. Se a alegação fosse uma com a qual não estamos familiarizados, i.e., se não soubéssemos de antemão seu valor de verdade, não consideraríamos a alegação explicativa por si. Ainda assim, Hempel amplia sua resposta. Para ele, alegações como (1.1) podem ser pensadas como contendo um argumento explicativo tácito que serve de *explanans* do qual poderíamos deduzir (1.1) como *explanandum*; neste argumento incluiríamos leis como as leis mecânicas sobre transferência de energia e leis sobre distribuição de moléculas em líquidos; e condições iniciais como a força necessária para derrubar o tinteiro, a espessura da tinta, o peso do tinteiro e sua aderência com a mesa.²

Proponentes do modelo ND tinham em mente certas concepções empiristas sobre a causação. Noções modais como “lei”, “causa”, “contrafactualmente suportado” formariam um conjunto de conceitos operativos para uma teoria. Por exemplo, o modelo ND precisa da noção de lei para explicitar o caráter científico de certas explicações. Noções modais não seriam claras por se referirem a qualidades disposicionais, que por isso não são atuais. Exemplificando, a diferença entre “branco” e “solúvel” é que podemos observar atualmente a brancura do sal, mas não sua solubilidade. A primeira qualidade é atual e a segunda é disposicional. Como disposições não são observáveis e envolvem certa obscuridade, a única maneira de esclarecermos qualidades modais seria fornecendo uma análise de termos modais em termos não modais. Se quiséssemos explicar noções modais nos referindo a outras noções modais teríamos uma circularidade não esclarecedora sobre essas noções.

Uma das consequências dessa concepção, para os defensores do modelo ND, é que alegações causais singulares, como (1.1), só podem ser explicadas por referência a leis que a suportem, mesmo que tacitamente. Uma das possíveis teses que decorrem do apelo a estruturas explicativas tácitas é que o *explanans* tácito implicaria a existência de uma lei geral

1 Iremos nos referir a algumas frases, exemplos e alegações usando o número da seção mais o número da ordem em que elas aparecem na seção.

2 Hempel fornece uma análise similar para lidar com as críticas sobre a inadequação ao modelo ND para as ciências especiais. Em particular, em seu (1942) mostra como poderíamos descrever a História como fornecendo explicações através de generalizações testáveis empiricamente, ou que se encontrem explícitas ou tácitas pela nossa familiaridade.

sobre o *explanandum*. No caso do *explanans* tácito de (1.1) haveria uma lei conectando impactos de joelho com quedas de tinteiros da forma “sempre que um impacto com tais e tais características ocorre, o tinteiro cai”, e seria essa última lei capaz de explicar porque (1.1) é explicativa. Mas, como Scriven havia notado, um homem das cavernas já sabia que (1.1) era verdadeira sem conhecimento de leis. Além disso, como a noção de lei surgiu tardiamente no pensamento humano, seria estranho dizer que não haviam explicações científicas anteriores ao estabelecimento da noção de lei – que inclusive ainda aguarda estabelecimento.

Supostamente, a noção de lei seria menos problemática que as outras noções modais porque temos casos paradigmáticos de leis. Isto nos leva a crítica externa ao modelo ND: como dar conta das ditas ciências especiais que não recorrem, ao menos não em todas explicações, às leis paradigmáticas? A lei da oferta e demanda não é uma generalização de escopo irrestrito como a lei da conservação da energia. Apesar disso, a lei da oferta e demanda parece explicar melhor fenômenos econômicos do que reduções físicas destes fenômenos. Se uma explicação que não atende ao critério de expectabilidade nômica não é científica, mas explica um fenômeno de maneira mais adequada do que uma explicação que atenderia ao critério de expectabilidade (por exemplo via redução dos fenômenos econômicos a fenômenos físicos), ficamos com um impasse: ou o modelo ND não se aplica à todas ciências e não pode ser o modelo explicativo de todas, ou devemos considerar as ciências especiais como ramos subdesenvolvidos de ciências melhores, como a Física e a Química.

Apesar das críticas, o modelo ND não teria centralizado a discussão sob modelos explicativos se fosse simplesmente inadequado, o modelo certamente captura certas intuições nossas sobre o que uma explicação deveria possuir. Woodward (2003, pp. 184-186) aponta três pontos positivos principais capturados pelo modelo ND e que devem ser buscados por quem queira fornecer uma proposta alternativa: i) Objetividade. O modelo não recorre a padrões subjetivos sobre como agentes cognitivos estariam mais ou menos dispostos a aceitar uma explicação como sendo boa; ii) Tratamento unificado. Ao mostrar como deve ser a relação entre *explanans* e *explanandum* o modelo ND fornece uma maneira de pensarmos em uma estrutura explicativa compartilhada por diferentes ciências. iii) Adequação com à ciência.

Nossa proposta de solução aos problemas do modelo ND, é começar dando um tratamento à noção de “causa”, e depois disto mostrar como uma teoria que comece utilizando a noção de causa pode fornecer modelos explicativos que poderiam ser utilizados nos casos em que o modelo ND falha.

A teoria que utilizaremos aqui é a teoria intervencionista da causalidade proposta em Woodward (2003). Ela se pronunciará sobre as noções modais mas não será reducionista; ela não pretenderá fornecer traduções redutíveis dos termos modais para termos não modais. Apesar disso, defendemos que ela pode ser circularmente esclarecedora ao fornecer um quadro conceitual compatível com nossas intuições e com a prática científica. Se a teoria que propusermos estiver correta teremos inclusive certas dúvidas quanto a uma análise reducionista ser esclarecedora. No capítulo 2 expomos a teoria causal e no capítulo 3 mostramos como as noções introduzidas no capítulo 2 podem resolver as críticas ao modelo ND.

CAPÍTULO 2 – A TEORIA INTERVENCIONISTA DA CAUSALIDADE

2.1 Localização da Teoria Intervencionista

A teoria intervencionista que utilizaremos é a proposta em Woodward (2003). Ela é uma teoria contrafactual da causalidade, mas não uma teoria reducionista da causação a contrafactuais. Teorias contrafactuais reducionistas da causalidade diriam que podemos expressar os conteúdos modais apelando apenas a contrafactuais – por exemplo Lewis (1984, 1994). A teoria intervencionista de Woodward também difere de teorias causais da agência, que pretendiam reduzir relações causais àquelas relações em que agentes pudessem ao alterar uma causa alterar seu efeito correspondente – por exemplo Price e Menzies (1993). Esquemáticamente, poderíamos representar a localização dessas últimas teorias assim:³

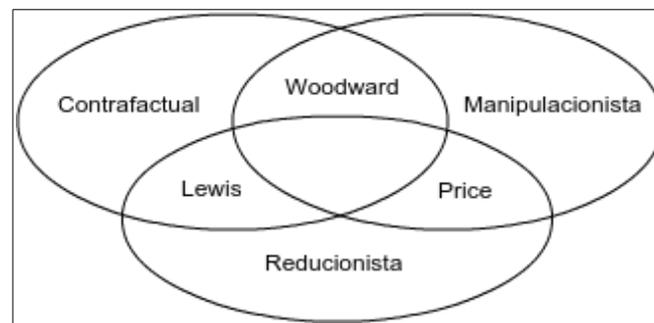


Figura 1: Panorama de Teorias Causais

Teorias causais da agência são reducionistas pois podemos substituir todos casos de “A causa B” por “um evento A é uma causa de um distinto evento B apenas no caso em que trazer a ocorrência de A seria um meio efetivo pelo qual um agente livre poderia trazer a ocorrência de B”. (MENZIES & PRICE, 1993) Como podemos notar, o *definiens* de “causa” não utiliza “causa”. Dado esse passo reducionista, é possível caracterizar qualquer causa como um caso de agência.

Não seria “trazer a ocorrência” uma forma escamoteada de dizer “causar”? Não para os defensores da teoria agencial, para os quais compreendemos ostensivamente noções causais, assim como podemos compreender ostensivamente uma cor. Por exemplo, assim como quando nos apresentam visualmente objetos verdes (folhas, grama, gafanhotos)

³ O quadro de teorias causais é mais amplo do que o exposto aqui. No entanto, as teorias localizadas aqui se aproximam bastante da intervencionista, e por isso nos restringiremos a essas. Para diferentes teorias causais ver Beebe, Hitchcock & Menzies (2009).

compreendemos o que significa atribuir “verde” a outros objetos da mesma cor; também compreendemos ostensivamente “trazer a ocorrência” pois somos agentes que realizam ações como meios para determinados fins. Se queremos obter B, e pensamos que realizar A aumenta a chance de B ocorrer mais do que não realizar nenhuma ação, realizamos A. Temos experiência direta desse tipo de caso e podemos entendê-lo sem uma concepção prévia do que seja causar, assim como podemos ter experiência direta de uma cor sem saber o nome daquela cor. A teoria agencial usa este caso de conhecimento direto, passível de mostragem ostensiva, como distintivo do nosso conhecimento sobre causalidade.

Embora as teorias causais de agência tenham perdido popularidade, elas foram as precursoras das teorias manipulacionistas e capturaram certos elementos importantes de relações causais, como as motivações pelas quais passamos a falar de causação (como exporemos em 2.3). O problema é trazer o papel do agente muito próximo para explicar o que a causação *é* e com isso possivelmente inviabilizar os casos de atribuição causal para os quais não podemos pensar em semelhança ostensiva com nossos casos de causação. Podemos pensar que astros causem alterações nas órbitas de outros astros, mas por que pensar que a nossa capacidade cognitiva deva ser relevante para determinar se temos uma relação causal?

A diferença crucial entre a teoria intervencionista e a agencial é que a primeira se compromete em fornecer uma teoria sobre o que é a causação, enquanto a segunda, pretende nos fornecer uma maneira de falarmos da causalidade. Como Price coloca, dado o comprometimento projetivista de pensarmos a causação como uma qualidade secundária “a melhor maneira de entendermos a causação não é perguntar o que a causação *é*, mas o que o discurso sobre a causação *faz* – qual função ele cumpre em nossas vidas.” (1992, p. 519, grifos originais.)

Claro que nossa capacidade cognitiva é importante para detectarmos relações causais, ninguém diria o contrário. Além disso, na seção 2.9 argumentaremos que só podemos nos posicionar sobre as relações causais, e das contrafactuais a elas associadas, das quais podemos saber o que ocorreria se as variáveis descritas pela relação tivessem seus valores alterados. Para a teoria intervencionista a causação *é* a associação entre a alteração do valor da causa e da alteração do valor do efeito (como caracterizaremos na seção 2.4), mas essa alteração *não é* tomada como um modo pelo qual viemos a falar da causação.

Dentre as teorias contrafactuais da causalidade a mais conhecida é a de Lewis, que pretende mostrar como relações causais como “*c* causa *e*” são redutíveis a contrafactuais da forma “se *c* não tivesse ocorrido, *e* não teria ocorrido”.

A proposta de Lewis para explicar o conjunto de conceitos modais está conectada ao seu projeto da sobreveniência humeana (*humean supervenience*) que afirma que “toda a verdade sobre um mundo como o nosso sobrevém na distribuição espaço-temporal das qualidades locais.” (1986. p.473), de forma charmosa podemos dizer que a verdade sobrevém aos entes. A ideia é mais intuitiva do que pode parecer. Tudo que existe são propriedades locais perfeitamente naturais que não precisam mais do que pontos em que possam ser instanciadas. Estes pontos formariam um mosaico do nosso mundo, e as verdades sobre o nosso mundo seriam sobrevenientes ao mosaico. A semelhança no mosaico acarreta semelhança na verdade, mas pode haver semelhança na verdade sem semelhança no mosaico. Além disso, só há diferença sobre certa verdade se houver diferença no mosaico (1994, p. 474). Um exemplo de sobreveniência pode tornar mais clara a posição lewisiana. O valor monetário é uma propriedade sobreveniente tanto de cédulas quanto de moedas. Cédulas de um real possuem o mesmo valor monetário, de forma que todas as cópias impressas possuem igual valor. O mesmo valor monetário (um real) é também encontrado em moedas de um real. A reprodução da cédula garante a reprodução do valor, mas podemos encontrar o mesmo valor sem a reprodução da cédula. Semelhança de nível inferior, moedas e cédulas, garante semelhança de nível superior, um real; porém semelhança de nível superior, um real, não garante semelhança de nível inferior, moeda ou cédula.

As leis naturais são as regularidades que podemos observar do mosaico (i.e., regularidades sobre o conjunto de propriedades locais que constituem nosso mundo) que incluiríamos em sistemas dedutivos que melhor combinassem força e simplicidade. (idem, p. 478-479) Dada a sobreveniência humeana, as leis que temos descrevem adequadamente o nosso mundo porque só seriam diferentes se o mundo fosse diferente. O mesmo que ocorre com a sobreveniência do valor monetário ocorre com a sobreveniência humeana. Assim como sabemos que as mesmas moedas só possuiriam valores monetários diferentes se as moedas fossem diferentes, também sabemos que determinada alegação nômica sobre nosso mundo é verdadeira porque ela só não seria verdadeira se o que observássemos fosse diferente.

A abordagem de Lewis foca na causação de eventos particulares (token) e não em tipo (type). Exemplificando, a alegação causal “a pedra lançada por João às 19:35 em 16 de Outubro de 1987 quebrou aquela garrafa” é um token, i.e., uma ocorrência particular, da alegação causal de tipo “pedradas quebram garrafas”. Podemos descrever a abordagem de Lewis como relacionando ocorrências de eventos, onde “ $O(x)$ ” indica a ocorrência do evento x e “ $\sim O(x)$ ” indica a não ocorrência de x . Para Lewis c causa e se existe uma cadeia de dependência contrafactual ligando c a e ($c_1...e_1, c_n...e_n$). Para caracterizarmos uma relação

como causal temos de atender dois critérios:

$$2.1.1 O(c) \rightarrow O(e)$$

$$2.1.2 \sim O(c) \rightarrow \sim O(e)$$

Na semântica de Lewis, a contrafactual (2.1.1) é automaticamente verdadeira – porque ela já descreve uma sequência de eventos ou correlação que é verdadeira. Consequentemente, o que é crucial na teoria de Lewis para e ser contrafactualmente dependente de c é que a contrafactual (2.1.2) seja verdadeira. Basicamente, a ideia é que todos os mundos em que (2.1.2) é verdadeira são mais próximos que os mundos onde o antecedente é verdadeiro e o consequente é falso. O critério de lei será crucial para avaliar a proximidade de mundos. Grosso modo, os mundos mais próximos ao nosso não devem incluir grandes violações de leis,⁴ que são os mundos que tornariam (2.1.2) falsa.

A teoria de Lewis conseguiu responder problemas que outras teorias regularistas sobre a causalção não tinham resolvido até então, o que contribui para seu prestígio. Ambas teorias, a intervencionista e a contrafactual de Lewis, consideram o papel das contrafactuais importantes e possuem pontos de convergência.⁵ A diferença crucial é que para a teoria intervencionista as únicas contrafactuais que são relevantes são aquelas que compreendemos o que é para as variáveis (como c e e) terem seus valores alterados – este será o tema da seção 2.8. O nosso comprometimento com a relação entre “dependência contrafactual” e “causação” é diferente, de modo que o último explica o primeiro e não o contrário. As dificuldades de explicar a causalção apelando apenas a dependência contrafactual envolvem os casos onde temos dependência contrafactual sem causalção, como casos de preempção (*preemption*) ou transitividade.⁶

4 Lewis (1986, p. 47-48) fornece os critérios na seguinte ordem de importância: “(1) It is of the first importance to avoid big, widespread, diverse violations of law. (2) It is of the second importance to maximize the spatio-temporal region throughout which perfect match of particular fact prevails. (3) It is of the third importance to avoid even small, localized simple violations of law. (4) It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly.”

5 Por exemplo, para Lewis devemos pensar as alterações nas causas como pequenos milagres cirúrgicos que alterariam apenas a causa e nada mais, enquanto em Woodward pensamos nas intervenções como alterando os valores das variáveis sem alterar outros valores da estrutura.

6 Preempção são casos onde C causa E e D não causa E , mas se C não tivesse ocorrido, D teria causado E . Se a causalção consiste apenas em dependência contrafactual temos o problema de explicar que C é uma causa de E porque a contrafactual “se C não tivesse ocorrido, E teria ocorrido” é verdadeira. O problema da transitividade é que qualquer evento pode acabar sendo caracterizado como causa de outro evento dada a história de dependência contrafactual contada. Trataremos a transitividade na seção 2.4.

2.2 Motivação para uma abordagem intervencionista

Por que uma teoria causal? E por que uma teoria manipulacionista da causalidade?

Existem diferentes interesses em relações causais que são refletidos nos mais diversos lugares em que atribuímos causalidade, quando nos perguntamos, por exemplo: se o medicamento que está em teste foi a causa da recuperação dos pacientes; se foi a negligência de um médico que causou a morte de um paciente; se determinado composto químico causa degradação ambiental; se a lua causa o aumento das marés. O nosso interesse em relações causais também impregna a vida cotidiana: queremos saber a causa do carro não ligar; a causa do bolo não crescer; o que causou o congestionamento; e assim por diante.

Em geral, quando queremos saber a causa de um efeito, queremos saber algo mais específico do que a mera sequência entre eventos, procuramos ver como a causa é relevante para o efeito; querer conhecer relações causais é querer saber mais do que correlações entre eventos.⁷ Possuir informação sobre relações causais é possuir informação que nos permita prever e, ocasionalmente, agir de forma mais adequada, por exemplo: saber se o composto químico é o responsável pela degradação ambiental, é saber se eliminá-lo evitará uma maior degradação; e saber que não havia fermento no bolo, nos ajuda a cozinhar melhor.

Como ilustramos acima, teorias causais podem responder a diferentes programas em diferentes áreas de conhecimento. No entanto, aqui usaremos “teoria causal” para as teorias que pretendem se aplicar a qualquer tipo de relação causal e responder o que há de comum por trás das diferentes alegações causais elencadas acima. Se conseguirmos um tratamento unificado é de se esperar que possamos voltar às diferentes alegações com algo que lhes possa ser benéfico.⁸ É o que a abordagem intervencionista pretende ao fornecer um arcabouço que pretende fazer três coisas: capturar nossas intuições sobre a causação; se adequar à prática científica; e ser utilizável em casos onde estamos incertos da existência de relações causais. Uma teoria causal que não atenda a esses requisitos pode ser considerada incompleta.

A teoria intervencionista que apresentaremos é a teoria intervencionista de Woodward (2003) que foi em grande medida inspirada pelo desenvolvimento anterior, liderado

7 Um regularista defenderia que relações causais são somente correlações, mas neste caso ele teria uma posição filosófica para defender, e deve explicar por que o senso comum distingue correlações de relações causais, de maneira que dizem que existe uma diferença entre “pedradas e vidros quebrados” e entre “o consumo per capita de margarina e a taxa de divórcios em Maine”. Para aquela e outras correlações, ver: <http://tylervigen.com/spurious-correlations>. Último acesso: 25 de Novembro de 2015.

8 Pode-se recusar que exista algo comum entre as diferentes alegações e considerar que o uso de “causa” em nossas teorias é mais danoso que benéfico – que é a visão de Russell em seu (1917).

principalmente por Judea Pearl e Clark Glymour, das redes de Bayes para sistemas de inteligência artificial que buscavam métodos matemáticos para assimilação de nova informação para tais sistemas (WALKER & GOPNIK, 2013, p. 345). Optamos pela abordagem de Woodward por tratar mais diretamente de problemas filosóficos.

Uma teoria intervencionista responde de maneira muito direta porque gostaríamos de possuir uma teoria causal: porque queremos saber se certas intervenções que fizermos alterará o resultado: queremos saber se adicionar fermento faz o bolo crescer; se remover determinado composto evitará degradação ambiental... Queremos saber se alterar uma coisa alterará outra, o que por muitas vezes possui implicações práticas. Analisemos os três modelos causais:

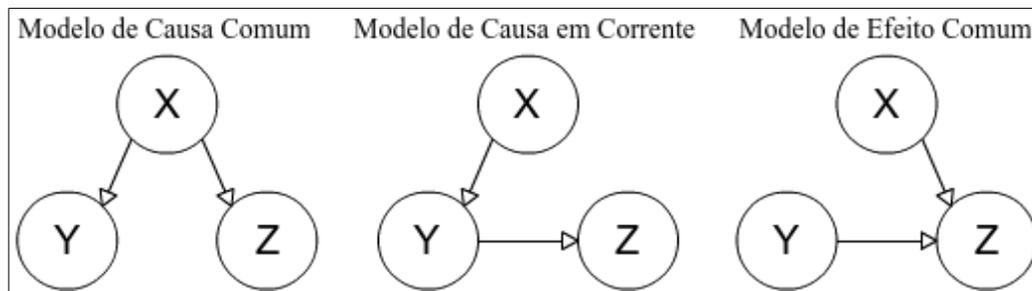


Figura 2: Modelos de representação causal. Baseado em Walker & Gopnik (2013, p. 346)

Também usaremos setas para expressar relações causais. Por exemplo, expressaremos as estruturas da Fig. 2 do seguinte modo: “ $Y \leftarrow X \rightarrow Z$ ” (X causa Y e X também causa Z); “ $X \rightarrow Y \rightarrow Z$ ” (X causa Y , e Y causa Z); e “ $X \rightarrow Z \leftarrow Y$ ” (X e Y causam Z). Se temos três variáveis que investigamos, escolher a estrutura correta pode ser relevante, por mais que a correlação entre as variáveis se mantenha sem a explicitação do conteúdo causal por trás da correlação. Existe correlação entre: fumar (X), dedos amarelados de nicotina (Y) e câncer de pulmão (Z), de tal modo que quase sempre acompanhamos a presença conjunta das três variáveis. Saber a estrutura causal permite-nos agir de maneira adequada para obtermos certo resultado. Se supomos que o amarelamento dos dedos pela nicotina causa o câncer de pulmão (pela estrutura $X \rightarrow Y \rightarrow Z$), então poderíamos desenvolver métodos para evitar o amarelamento, e conseqüentemente o câncer pulmonar, e continuar fumando sem risco de desenvolver câncer pulmonar. Por outro lado, se tomamos o modelo de causa comum ($Y \leftarrow X \rightarrow Z$), alterar o amarelamento dos dedos não diminuirá a probabilidade de desenvolvermos câncer pulmonar, e com isso não procuraríamos meios de evitar o amarelamento com fins de evitar o câncer.

Nem sempre será trivial escolhermos um modelo adequado, mas o nosso interesse em descobri-lo sugere que nosso interesse vai além de descobrir apenas a correlação entre as

variáveis. Pense-se nas correlações: entre adquirir seguro de vida e possuir uma expectativa de vida maior; ou entre estudar em escolas privadas e apresentar maiores sucessos acadêmicos. Em ambos os casos saber se uma coisa causa a outra nos permite desenhar diferentes testes para testar hipóteses e agir de acordo. Embora possa ser inadequado fechar todas escolas públicas e custear ensino privado aos contribuintes de um país, descobrir a relação adequada entre as variáveis (inclusive as não explícitas na relação, como condição de saúde dos alunos) permite agir para melhorar globalmente os sucessos acadêmicos.

Dizer que o nosso interesse em relações causais é proveniente apenas de prazer intelectual desinteressado pragmaticamente nos joga o ônus de explicar como animais não-humanos e crianças já se interessam por relações causais. Como Woodward aponta (2007, p. 32), embora primatas não possuam conhecimento causal acurado sobre relações não-egocêntricas,⁹ eles reconhecem-se como agentes causais, e aprendem que podem replicar ações que observam feitas por outros seres que reconheçam como agentes. Walker e Gopnik (2013) mostram como a imaginação de crianças já lhes fornecem um vocabulário que consegue distinguir entre os modelos de causa comum, efeito comum e causa em corrente. Gopnik et. al. (2004) mostram que crianças já possuem um aparato causal capaz de prever resultados de como intervenções trariam certos resultados, mesmo que não tenham experimentado os resultados (idem, pp. 24-29), o que sugere que elas já pensam contrafactualmente. Consideramos um ônus para quem defende uma concepção extremamente intelectualizada da causação explicar como esses casos não refletem conhecimento causal.

Resumidamente, diríamos que uma abordagem intervencionista é motivada por mostrar uma maneira gradativa de explicar como pensamos nas relações causais desde a infância até a vida adulta, e também dos contextos ordinários aos científicos. Embora Woodward (2003) forneça uma maneira de compreendermos a qualidade de uma explicação através do grau de invariância de uma explicação (seção 2.7), ele não fornece uma história de como podemos adquirir um vocabulário causal, e talvez isto torne desconfortável uma abordagem circular da

9 Woodward (2007, p. 32), baseado em Tomasello e Call (1997), toma deles a divisão entre níveis de conhecimento causal, onde os três argumentam que primatas não-humanos não possuiriam o nível de conhecimento causal mais refinado que envolveria reconhecer relações causais terciárias, onde as relações se dão entre o que os primatas não reconhecem como agentes. Os níveis, de menor a maior refinamento, são os seguintes: “1. An agent whose instrumental behavior and learning is purely egocentric. That is, the agent grasps (or behaves as if it grasps) that there are regular, stable relationships between its manipulations and various downstream effects but stops at this point, not recognizing (or behaving as though it recognizes) that the same relationship can be present even when it does not act, but other agents act similarly or when a similar relationship occurs in nature without the involvement of any agents at all. 2. An agent with an agent causal viewpoint: The agent grasps that the same relationship that it exploits in intervening also can be present when other agents act. 3. An agent with a fully causal viewpoint: The agent grasps that the same relationship that the agent exploits in intervening also can be present both when other agents intervene and in nature even when no other agents are involved. This involves thinking of causation as a tertiary relationship.”

causação. Contudo, o trabalho liderado por Alison Gopnik (GOPNIK, GLYMOUR, SOBEL, KUSHNIR E DANKS, 2004) fornece uma análise que é completamente compatível com a teoria causal que desenvolveremos, e que pretende dar conta do aprendizado causal em crianças. O final desta seção é um *esboço* da proposta de Gopnik com o intento de tornar mais assimilável como podemos adquirir conhecimento de base das relações causais.

O trabalho liderado pela Gopnik aplica redes bayesianas, que são equações e gráficos associados para tais equações (que utilizaremos na seção 2.3) com inclusão de certos aspetos formais (em Gopnik, et. al., 2004, pp. 8-18) para mostrar como bebês e crianças aprendem a pensar causalmente. A hipótese é que pelas probabilidades que as crianças associam ao observar sequências de eventos elas constroem um mapa causal do mundo, o mapa então é usado para fazer novas inferências e interpretar possíveis conteúdos causais. As redes bayesianas permitem quantificar como ocorre o desenvolvimento desse mapa, i.e., quantificar como se dá a alteração da probabilidade com a qual elas esperam que uma sequência de eventos se mantenha, ainda que obviamente as crianças não saibam explicitar isso estatisticamente. Para Gopnik, assim como para Woodward, intervenções constituem a maneira que mais aumenta a probabilidade de tomarmos uma sequência como sendo causal¹⁰, pois isolamos as outras variáveis que poderiam estar causando uma correlação. Pode ser que surja um certo verificacionismo da posição de Woodward, e que parecerá reforçada na seção 2.8 onde argumentaremos que só podemos falar das relações causais das quais temos conhecimento do que seria para as variáveis de uma relação causal terem seus valores alterados. Ainda assim, ambos afirmam que podemos extrair conteúdo causal de correlações. O trabalho da Gopnik permite capturar a passagem que crianças fazem de A) para B):

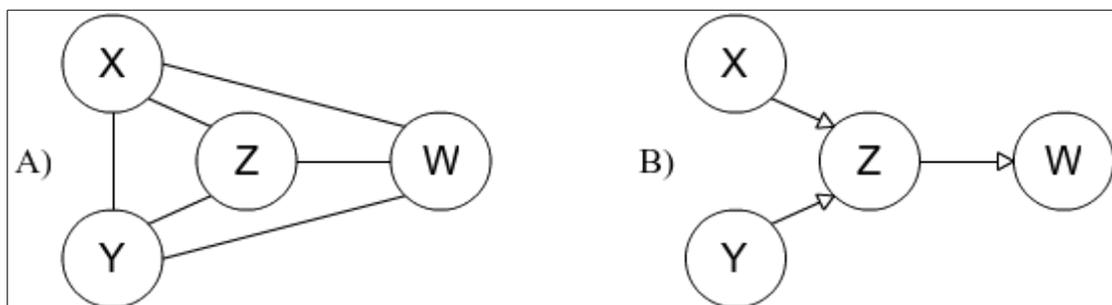


Figura 3: Transição entre estruturas causais. Baseado em Gopnik et. al. (2004, pp. 14-16)

Onde na estrutura A) teríamos apenas correlações, e em B) já possuiríamos as relações

¹⁰ Gopnik et. al. (2004, pp. 24-29) testaram essa hipótese através da “Puppet Machine” que consistia de uma caixa com um mecanismo de causa comum oculto às crianças. Nos dois testes onde as crianças observaram intervenções elas acertaram mais estrutura causal (78% e 84%) do que as as crianças do grupo de controle que não observaram intervenções (31% e 34%).

causais que forneceriam um mapa causal que serviria de conhecimento de fundo para analisar outras relações causais. Embora comecemos com associações probabilísticas entre eventos, no final das contas, o que precisamos para que uma relação seja causal é que a alteração do valor da causa, acarretará alterações no valor do efeito, e é sobre isso que argumentaremos no restante deste capítulo.

2.3 Apresentação Informal da Teoria

Começamos com uma abordagem informal da teoria que pretende facilitar as caracterizações mais detalhadas que faremos depois. A ideia por trás da teoria causal é que as relações causais são relações funcionais entre variáveis; e X causa Y , quando ao intervir em alterar o valor de X , o valor de Y será alterado. Mais precisamente:

A noção de uma intervenção representa uma tentativa de capturar, em uma linguagem não-antropomórfica que não faz referências a noções como agência humana, as condições que precisariam ser atingidas em uma manipulação experimental ideal no valor de alguma variável X realizada com propósito de determinar se X causa uma segunda variável Y . Mais precisamente, uma intervenção em X (com respeito a Y) é um processo causal que diretamente altera o valor de X de tal maneira que, se uma alteração no valor de Y ocorresse, ocorreria somente pela alteração no valor de X e não de outra forma. (WOODWARD, 2000, p. 4)

Temos uma relação invariante quando podemos descrever como se dá uma relação funcional entre variáveis dentro de um sistema. De maneira esquemática, para termos uma relação causal precisamos de uma função do tipo $Y=F(X_1...X_n)$ onde F rastreie como diversas causas ($X_1...X_n$) estão associadas a um efeito (Y). No caso mínimo, teremos dois valores x_1 e x_2 , e dois valores y_1 e y_2 , tais que $x_1 \neq x_2$ e $y_1 \neq y_2$, onde $y_1 = F(x_1) \neq y_2 = F(x_2)$. Invariância é uma propriedade gradativa, mais tarde (seção 2.7) caracterizaremos melhor como capturar o grau de invariância.

As equações estruturais são equações que permitem a construção de modelos representacionais.¹¹ Isto é feito através de um par ordenado $\langle V, E \rangle$, onde V é um conjunto de

¹¹ Redes bayesianas, como mencionamos na seção 2.2, são equações estruturais com gráficos para as equações. Embora as redes bayesianas possam ser usadas com concepções diferentes de causalidade ou sem qualquer conteúdo causal, usaremos sempre a concepção intervencionista da causalidade quando usarmos os gráficos. Encontramos variações de nomenclatura na literatura: a estrutura pode ser chamada de rede ou modelo; as variáveis podem ser chamadas de elos ou nós da rede; as setas, que indicam rotas causais, podem ser chamadas de vértices; e um conjunto qualquer de nós ligados por setas pode ser chamado de cadeia ou corrente. Em uma relação entre nós ligados por setas, como " $X \rightarrow Y$ ", dizemos que X é pai ou antecedente de Y , e que Y é filho ou descendente de X . Usamos caixa alta para variáveis e caixa baixa para os valores assumidos pelas variáveis, e para ambas itálico. Variáveis que designam conjuntos estão em negrito.

variáveis e E um conjunto de equações para as variáveis. Há variáveis endógenas e variáveis exógenas. As variáveis endógenas são as variáveis que têm seus valores determinados unicamente pelas equações internas do modelo. As variáveis exógenas têm seus valores determinados por fatores externos ao modelo. Na medida em que não possuem variáveis determinando seu valor, as variáveis exógenas funcionam como dados de entrada para o modelo explicativo. As equações estruturais permitem a construção de gráficos direcionados que ilustram o direcionamento das variáveis, isto é, em que direção ocorre a causação.

Para ilustrarmos a noção de intervenção, assim como das equações estruturais, usaremos um rádio.¹² Nosso rádio tem duas teclas e três botões. L é uma tecla liga/desliga. B é uma tecla que altera entre AM e FM. Os três botões, dials que giram 180° (onde 0° é valor mínimo e 180° o valor máximo), são: $D1$ e $D2$, respectivamente para AM e FM, onde a cada posição do dial corresponde uma captação da antena; e $D3$ é o dial do volume. Por fim, S é o alto-falante de saída. O sinal de saída é determinado pelo sinal que ele recebe, ou de $D1$ ou de $D2$, e é amplificado por $D3$. Contrariamente aos rádios modernos, o botão de volume não habilita a saída de som, simplesmente amplifica o sinal de saída. As equações estruturais para o nosso rádio são as seguintes:

$L =$ ligado ou desligado.

$B =$ Se $L =$ ligado, então AM ou FM.

$D1 =$ Se $B =$ AM, então 0° a 180° correspondendo a captação da antena.

$D2 =$ Se $B =$ FM, então 0° a 180° correspondendo a captação da antena.

$D3 =$ Se $L =$ ligado, então 0° a 180° correspondendo ao volume mínimo e máximo.

$S = D3 + (D1 \text{ ou } D2)$

Do lado esquerdo das igualdades temos o conjunto V , e do lado direito o conjunto E . O gráfico direcionado associado as equações é o seguinte:

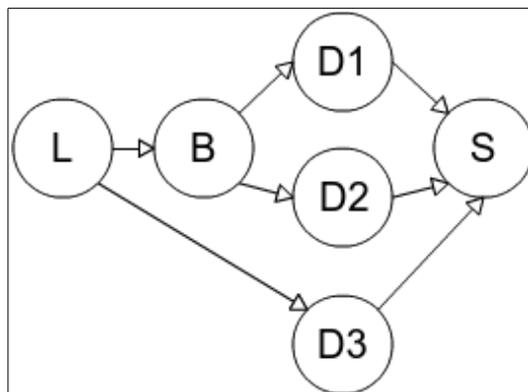


Figura 4: Gráfico direcionado de rádio

¹² Woodward & Hitchcock (2003, p. 12) já usavam um exemplo com rádio, o nosso é inspirado naquele.

Na estrutura da Fig. 4 a única variável exógena é L e todas outras são endógenas pois parte da determinação de seu valor depende de outras variáveis; graficamente isso é representado por L não ter setas direcionadas a si, enquanto as outras variáveis possuem. Em uma intervenção precisamos que a alteração do valor da causa ocorra de maneira a ser independente dos outros valores. É como se a relação manipulada isolasse sua relação com as outras variáveis.¹³ Não é necessário que ela efetivamente seja isolada, mas que a alteração do valor da causa que alterará o valor do efeito não seja influenciada por esses valores. Podemos notar que DI será uma causa de S , pois quando L =ligado, B =AM e $D3$ fixo em qualquer valor, a alteração da posição de DI altera S .

Uma intervenção é algo que altera o valor da variável da causa para alterar o valor do efeito. A intervenção é uma variável exógena ao modelo. Sendo I uma variável que representa a intervenção de girar o dial ficamos com a seguinte estrutura causal:

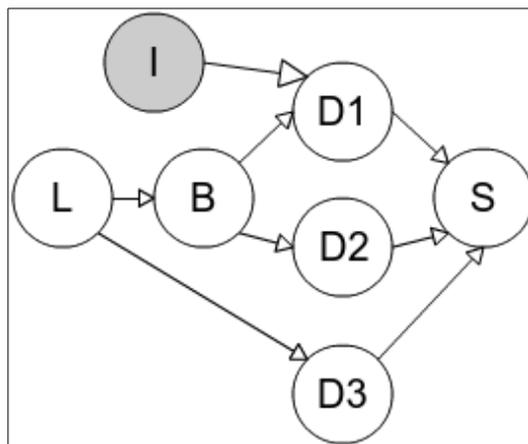


Figura 5: Variável Intervindo no Rádio

A relação causal $I \rightarrow DI \rightarrow S$ pode ser considerada isolada dos outros valores quando os outros valores que alteram S são fixados de maneira a não perturbar a relação causal. É quando conseguimos este tipo de relação causal que temos uma relação invariante sob manipulações. É importante apenas que para algumas alterações no valor da causa ocorram algumas alterações no valor do efeito, mas não que para toda alteração no valor da causa ocorra uma alteração no valor do efeito. Nesse exemplo se poderia pensar em posições em que DI não capta sinal de nenhuma estação, onde alterações entre essas posições não causam alterações em S . Outros exemplos são fáceis de se ver: aumento de temperatura causa derretimento de sorvete, mas não o aumento de qualquer temperatura para qualquer outra.

¹³ Falamos aqui das variáveis já descritas na estrutura, mas existem variáveis externas que romperiam as relações: em locais isolados da Terra onde não houvesse captação de rádio, ou fonte de energia, nenhuma alteração nos botões alteraria S . Na seção 2.7 tratamos detalhadamente dos pontos de ruptura de estruturas.

Quando tivermos alguns valores externos a relação $I \rightarrow DI \rightarrow S$ (como no exemplo $L = \text{desligado}$ ou $B = \text{FM}$) não haverá qualquer intervenção em DI que altere S . Mas para a caracterização de relação causal dentro da abordagem manipulacionista a existência de uma função associativa entre variáveis para *alguns* valores fixos, outros que os da própria relação causal (aqui $DI \rightarrow S$), já é suficiente para caracterizarmos a relação como causal.

Como podemos ler uma relação funcional entre as variáveis do rádio como uma relação causal? Quando podemos descrever que alterações de certas variáveis alteram o valor de outras variáveis. Supondo que o rádio esteja em Porto Alegre, teríamos as seguintes relações funcionais:¹⁴

$$F(DI=0^\circ) = \text{Real AM}$$

$$F(DI=34^\circ) = \text{Sem captação}$$

$$F(DI=91^\circ) = \text{Pampa AM}$$

$$F(DI=114^\circ) = \text{UFRGS AM}$$

$$F(DI=173^\circ) = \text{Sem captação}$$

Onde para a cada uma corresponderia respectivamente a contrafactual:

Se DI estivesse na posição 0° , S reproduziria a estação Real AM.

Se DI estivesse na posição 34° , S não reproduziria uma estação.

Se DI estivesse na posição 91° , S reproduziria a estação Pampa AM.

Se DI estivesse na posição 114° , S reproduziria a estação UFRGS AM.

Se DI estivesse na posição 173° , S não reproduziria uma estação.

Mais tarde (na seção 2.8) detalharemos mais a relação entre contrafactuais e intervenções. Podemos notar que existe uma correlação direta entre contrafactuais e a relação funcional descrita pela intervenção de alterar a posição de DI e a alteração em S .

Podemos notar que o exemplo do rádio captura a exigência mínima com a qual começamos esta seção, temos mais do que dois valores de causas x_1 e x_2 , e mais do que dois valores de efeitos y_1 e y_2 , tais que se alterássemos as causas, passando de x_1 para x_2 , teríamos uma alteração correspondente de y_1 para y_2 , e os valores assumidos pelas variáveis através da função associativa seria diferente, tal que $y_1 = F(x_1) \neq y_2 = F(x_2)$.

Através de uma linguagem intervencionista, ficamos com uma maneira intuitiva de tratarmos contrafactuais e entender qual seria seu suporte. Como? Descrevendo a relação funcional, dentro de um determinado sistema, de como a alteração do antecedente acarreta a alteração do conseqüente. Quando uma contrafactual é falsa? Quando não se mantém uma relação de alteração do antecedente para a alteração do conseqüente. Esquemáticamente, se

¹⁴ A tabela mostrando todas captações hipotéticas do rádio está no Apêndice A.

em uma relação destacada de sua estrutura: alterar x_1 para x_2 não altera y_1 para y_2 ; ou y_1 é alterado para y_2 sem que haja uma alteração de x_1 para x_2 . Isso é observado no rádio, se ao isolarmos $I \rightarrow DI \rightarrow S$ do sistema: alterações na posição de DI , através de I , não alteram S – giramos o botão e a estação não troca; ou se S é alterada independentemente de DI – a estação troca sem girarmos o botão. Se uma dessas duas coisas ocorrem, diríamos que o nosso rádio está com problemas e não temos mais uma relação causal, de forma que a lista de contrafactuais que elencamos anteriormente não seriam mais verdadeiras.

Há duas formas pelas quais uma relação funcional pode ser rompida, quando a alteração dos valores é muito maior do que a descrita na relação ou se a estrutura não inclui uma variável que poderia romper a estrutura. No nosso rádio, girar qualquer dial acima de 180° quebra o dial, e a relação entre dial e saída é rompida. Na estrutura também não incluímos variáveis para o local e o ano em que o rádio opera, alterar qualquer uma dessas variáveis mudaria as contrafactuais decorrentes da relação funcional que descrevemos entre DI e S , por exemplo, o rádio se estivesse na China não capturaria as estações, e em 1930 não captaria a UFRGS AM, que começou sua transmissão em 1957. Quando tratarmos a noção de invariância (na seção 2.8) daremos um tratamento mais detalhado destes rompimentos de estruturas causais, por ora, voltemos às intervenções.

Embora a teoria manipulacionista não precise recorrer aos gráficos direcionados e às equações estruturais, os dois juntos explicitam de maneira ilustrativa diversos pontos da abordagem. As equações e os gráficos são úteis para a teoria manipulacionista porque a noção central é a de causa direta. Quando X causa Y , X é uma causa direta de Y quando a alteração de X altera sem causas intermediárias o valor de Y . Não haver intermediários é facilmente detectado graficamente ou pelas equações – e na seção 2.4 caracterizaremos causa direta.

As equações são sempre mais precisas que os gráficos porque: i) gráficos iguais podem expressar relações causais diferentes entre as variáveis;¹⁵ e ii) equações expressam o que realmente precisa ocorrer para que uma variável altere outras variáveis. Por sua vez, os gráficos sempre são mais ilustrativos para mostrar panoramicamente o que causa o que dentro da estrutura causal, de forma que o uso conjunto é bastante útil. As definições de diferentes tipos de causa (discutidas na seção 2.4, e contidas em Woodward (2003, capítulo 2)) são independentes do modo como as relações causais são representadas, o que dá bastante margem para a utilização da teoria. É comum ver as equações estruturais associadas com

15 Por exemplo, poderíamos representar uma causa comum a dois efeitos, quando X causa Y e Z , igualmente por “ $Y \leftarrow X \rightarrow Z$ ” e usá-la para expressar as duas relações causais distintas: uma bola de bilhar batendo simultaneamente em duas e pondo-as em movimento; aumento dos preços de um produto, fazendo consumidores comprarem menos o produto e mais o do concorrente. Somente o gráfico não revela qual é a relação entre as variáveis.

relações lógicas e os valores $\{0, 1\}$ representando a ocorrência ou não de eventos¹⁶. Não precisamos assumir nem que a causalção seja uma relação entre eventos e nem que os valores sejam binários, embora possamos incorporar ambas. O problema de limitar a causalção a casos binários é que vão haver casos difíceis onde a insistência em dois valores poderá atrapalhar.¹⁷

2.4 A Teoria causal

Passemos agora aos detalhes da teoria causal. A versão mais formal da teoria é exposta em Woodward (2003, capítulo 2) com as definições para a causalção: condição necessária; condição suficiente; e de alteração de uma variável. A condição suficiente:

(CS) Se (i) há uma intervenção possível que altera o valor de X tal que (ii) realizar esta intervenção (e nenhuma outra intervenção) alterará o valor de Y , ou a probabilidade distributiva de Y , então X causa Y . (2003, p. 45)

Paralelamente, a condição necessária:

(CN) Se X causa Y , então (i) há uma intervenção possível que altera o valor de X tal que (ii) se esta intervenção (e nenhuma outra intervenção) fosse realizada, então o valor de Y (ou a probabilidade distributiva de algum valor de Y) alteraria. (idem.)

E de maneira próxima a como usamos anteriormente, alterar uma variável:

Se pensarmos em uma intervenção como tomando um valor particular por uma variável I , então I altera o valor de X e, ao fazê-lo, altera o valor de Y se e somente se há outros valores, x_1, x_2 , de X que $x_1 \neq x_2$ tal que para I assumindo algum valor i_1

16 Por exemplo Menzies (2014).

17 Por exemplo, em certos casos de preempção sobreposta (“trumping preemption”). Sobre discussão e exemplos, como o que usaremos aqui, ver Schaffer (2000). Em seguida acompanhamos de perto a exposição de Woodward (2003, p. 81-82) sobre o exemplo de Schaffer (2000, p. 175). Uma tropa segue as ordens de um major e de um sargento. As ordens do major sobrepõem-se as do sargento, no sentido em que a tropa faz o que o major manda independentemente das ordens do sargento. Mas, quando o major não dá ordens, a tropa segue as ordens do sargento. Se ambos mandassem a tropa avançar ela avançaria, e diríamos que a ordem do major causou o movimento da tropa, mas a do sargento não. Também diríamos que se o major não tivesse ordenado, a tropa ainda se moveria. Temos causalção, mas não dependência contrafactual da ordem do major para o movimento da tropa. Podemos resolver este caso usando variáveis que assumem mais de dois valores. Sejam os valores: -1= recuar; 0= não dar ordens; e 1= avançar. E as variáveis: M = ordem do major; S = ordem do sargento; T = ação da tropa. As equações seriam as seguintes:

Se $M \neq 0$, então $T = M$

Se $M = 0$, então $T = S$

A teoria intervencionista fornece uma maneira de capturarmos a intuição de que se $M=1$, não há intervenções em S que alterem T , mesmo que S possa causar T sob algumas manipulações. Embora a discussão de Schaffer seja mais sobre a inadequação de abordagens contrafactuais para a preempção sobreposta, uma representação gráfica do que ocorre já desambiguaria parte do problema levantado por ele.

causa X a assumir o valor x_1 e que se I assumindo algum valor i_2 causa X a assumir o valor x_2 e há os valores y_1, y_2 de Y associados com x_1 e x_2 tal que $y_1 \neq y_2$. Em outras palavras, se $Y=F(X)$ descreve a relação funcional entre X e Y , há valores distintos de X, x_1 e x_2 e valores Y, y_1 e y_2 tal que $y_1=F(x_1) \neq y_2=F(x_2)$. (idem)

Existe assimetria entre os lados da igualdade de $Y=F(X)$ que captura a ideia de que as causas determinam os efeitos, e não o oposto. Decorrente desta assimetria, o direcionamento das contrafactuais só se dá com o antecedente do lado esquerdo da equação e com o conseqüente do lado direito. São válidas as contrafactuais do tipo “se X houvesse ocorrido, então Y teria ocorrido” e não válidas as do tipo “se Y houvesse ocorrido, então X teria ocorrido”. Isto ocorre pois, dado o acontecimento do efeito, não podemos afirmar o acontecimento da causa. Essa assimetria podia ser observada no caso do rádio onde diversas posições de DI não possuem captação, onde saber que não temos captação não informa a posição de DI (ou sequer se o rádio está ligado), embora para saídas específicas de S pudéssemos saber a posição de DI .

$D3$ estava no exemplo do rádio para indicar que se não for controlado não teríamos certeza que é o próprio DI que é responsável pela alteração em S . Se supuséssemos uma ligação interna aos dials, de modo que ao girarmos DI giraríamos $D3$ conjuntamente, não poderíamos determinar qual é exatamente a estrutura causal que conecta os botões do rádio à saída de som do alto-falante. No rádio isso é trivial, mas em certas estruturas causais precisamos de mecanismos que rompam uma estrutura causal para verificarmos se as relações causais individuais estão presentes. Isso vai envolver uma distinção entre causalidade total e causalidade contribuinte. A distinção fica fácil de ver com um exemplo. No exemplo, o uso de pílulas anticoncepcionais (PA) é uma causa direta do aumento da probabilidade do surgimento de trombose (ST), e também diretamente causa a diminuição da probabilidade de gravidez (PG) que diretamente aumenta a probabilidade de trombose. Onde por fim, todos os efeitos de PA em ST são anulados pela rota de PA que passa por PG :

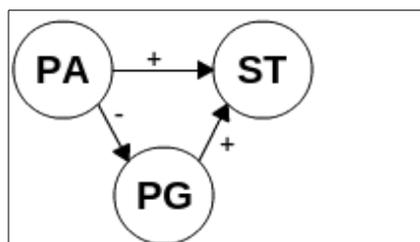


Figura 6: Gráfico PA, PG, ST

Se quisermos afirmar que PA tem algum efeito em ST , mesmo que não altere sua

probabilidade devido a rota causal $PA \rightarrow PG \rightarrow ST$, precisamos distinguir entre dois tipos de causalidade, a direta e a contribuinte:

Digamos que X é uma *causa total* de Y se e somente se há alguma intervenção apenas em X (e em nenhuma outra variável) tal que para algum valor de outras variáveis além de X , esta intervenção em X alterará o valor de Y . O *efeito total* da alteração Δx de X em Y é então, a alteração no valor de Y , ou na probabilidade distributiva de Y , que resultaria de uma intervenção apenas em X que altera seu valor pela quantidade Δx , dado os valores de outras variáveis. (Não é assumido que a noção de intervenção seja caracterizada de tal maneira que os valores de outras variáveis que não são descendentes de X não são alterados por uma intervenção em X). [...] A noção de causa total contrasta com a noção de *causa contribuinte*, a qual pretende capturar a ideia intuitiva de X influenciando Y por uma rota mesmo se, devido a cancelamento, X não possui efeito total em Y . [...] X será uma causa contribuinte de Y se e somente se ele [X] fizer uma contribuição não-nula em Y junto a alguma rota direta no sentido em que para aquelas variáveis (se alguma) que *não* estão nesta rota, há algum conjunto de valores para aquelas variáveis tal que se aquelas variáveis fossem fixadas por intervenções naqueles valores, há alguma intervenção em X que alterará o valor de Y . (WOODWARD, 2003, p.50, grifos do autor)

Como podemos observar, embora PA não seja uma causa total de ST , ainda podemos descrevê-la como uma causa contribuinte. Se pudermos romper a estrutura causal $PA \rightarrow PG$, ou analisá-la independentemente, poderemos verificar que PA é uma causa contribuinte de ST , mesmo que factualmente isso não ocorra. Isto poderia ser feito com dois grupos de mulheres estéreis, com um grupo tomando pílulas anticoncepcionais e outro não. Neste caso, se PA for uma causa contribuinte de ST , podemos esperar que a incidência de trombose no grupo que tomou as pílulas será maior do que a incidência no grupo que não tomou. Em outras palavras, embora a variação de PA (Δx) não cause efeitos totais em ST , pois PA nunca esteve factualmente dissociado de PG (e sem intervenções talvez nunca estará), se manipulássemos PG para mantê-lo fixo, poderíamos determinar que PA causa (contributivamente) ST .¹⁸

Embora CS e CN sejam adequadas para tratar da causalidade total, elas não o são para a causalidade contribuinte. Para a causalidade contribuinte precisamos reformular as condições necessárias e suficientes, e precisamos explicitar o que é para uma causa ser direta. O problema surge pois precisamos encarar sistemas que possuem variáveis interconectadas que podem cancelar a condição mínima de causação,¹⁹ mesmo que tal condição fosse atendida via intervenções nas variáveis fora da relação – que é o caso da Fig. 6.

18 O exemplo que usamos é de Hesslow (1976) apud Woodward (2003, pp. 49-51). Seguimos proximamente Woodward nesta exposição. Há um contraexemplo similar às equações estruturais em Menzies (2014), cuja solução é similar a adotada aqui. Naquele contraexemplo um assassino coloca veneno no café do rei (V), caso haja veneno, um guarda responde colocando antídoto no café do rei (A), o rei toma o café e sobrevive (S). Se possuímos apenas a noção de efeito total, o veneno não é uma causa da morte do rei: pois A está presente quando V está, e então temos S ; e A não está presente quando V não está, e novamente temos S .

19 De que existam dois valores para uma causa X , x_1 e x_2 , e dois valores para um efeito Y , y_1 e y_2 , tais que $x_1 \neq x_2$ e $y_1 \neq y_2$, onde $y_1 = F(x_1) \neq y_2 = F(x_2)$.

Para explicarmos a diferença entre causalidade total e contribuinte precisamos explicar a causalidade direta. Graficamente ela é observada quando existe apenas uma seta (\rightarrow) entre duas variáveis. Mas isto é pouco preciso, na definição de Woodward:

(CD) Uma condição necessária e suficiente para X ser uma causa direta de Y , com respeito a algum conjunto de variáveis V , é que exista uma possível intervenção em X que altere Y (ou a probabilidade distributiva de Y) quando todas outras variáveis em V , fora X e Y , estão fixadas em algum valor por intervenções. (2003, p. 55)

O conjunto V é relativizável, e com isso, uma variável ser uma causa direta é algo sempre relativo ao sistema. Porém, dada a determinação das variáveis, o resultado esperado pelas alterações das variáveis é objetivo. Poderíamos ter incluído no nosso rádio variáveis correspondendo a antena estar ligada, ao rádio estar conectado a tomada, ou ainda do detalhamento de como ele opera eletronicamente. Esses diferentes conjuntos de variáveis não correspondem a uma diferença nas relações causais objetivas do mundo, o funcionamento real do rádio é objetivo a despeito dessa escolha de variáveis. O que ocorre é que quando formulamos um modelo explicativo, escolhemos as variáveis que consideramos relevantes. Alguém que se interesse em consertar o rádio achará relevante incluir em seu modelo as variáveis com respeito ao funcionamento eletrônico, enquanto alguém que quer apenas sintonizar na UFRGS AM não achará; não se segue desse diferente conjunto de variáveis que o funcionamento da rádio deixa de ser objetivo.

Intimamente ligado à noção de causa direta está nossa propensão em poder pensar mecanismos causais como sendo distintos. Como pensamos que $PG \rightarrow ST$ opera de maneira diferente de $PA \rightarrow PG \rightarrow ST$, pensamos que elas podem ser expressas por relações diretas diferentes. No exemplo de Woodward (2003, pp. 55-56) podemos comparar duas estruturas: (2.4.1) “O desejo de vingança de A \rightarrow o aperto do gatilho \rightarrow a morte de B” e (2.4.2) “O desejo de vingança de A \rightarrow o aperto do gatilho \rightarrow a liberação da mola \rightarrow o impacto do martelo com o cartucho \rightarrow a explosão do cartucho \rightarrow a propulsão da bala para fora da arma em uma rota de impacto com o coração de B \rightarrow a morte de B”. Apertar o gatilho só é uma causa direta da morte de B relativamente a (2.4.1), mas se estivéssemos descrevendo um caso real, a morte de B não seria subjetiva.

Com a noção de causa direta podemos definir uma condição necessária para a causalidade contribuinte:

(NC*) Se X é uma causa contribuinte em nível de Y , com respeito a algum conjunto de variáveis V , então há alguma rota direcionada de X para Y , tal que para cada elo

nesta rota há uma relação causal direta; isto é, há variáveis intermediárias ao longo dessa rota, $Z_1...Z_n$, tais que X é uma causa direta de Z_1 , que é uma causa direta de Z_2 , que é uma causa direta de... Z_n . Posto diferentemente, se X causa Y , então X deve: ou ser uma causa direta de Y ; ou deve haver alguma cadeia causal, onde cada elo envolva uma relação de causalidade direta, estendida de X para Y . (2003, p. 57)

Como Woodward nota em seguida, se substituíssemos “se” por “se e somente se” adicionaríamos uma condição suficiente, mas ficaríamos com o problema da transitividade. Suponhamos um operador A que é destro e a seguinte cadeia causal: cachorro morde a mão direita de $A \rightarrow$ no dia seguinte A altera DI com a mão esquerda \rightarrow o rádio é sintonizado na UFRGS AM. Temos uma cadeia de dependência contrafactual, mas não estamos inclinados a dizer que a mordida causou a sintonização, muito embora a mordida tenha causado a sintonização com a mão esquerda. Como resolvemos esse problema?

Intuitivamente, pensamos que a mordida não causa a alteração da sintonia porque é uma informação irrelevante para DI . Se incluíssemos uma variável com os dois valores {usar mão direita, usar mão esquerda} não diríamos que a alteração entre esses valores é causalmente relevante para a alteração de DI . Quando pensamos em uma estrutura que contenha rotas não explícitas que romperiam a relevância da rota de X para Y , fica fácil vermos por que não estaríamos dispostos a atribuir a causalidade entre X e Y . Quando nosso interesse é saber em qual estação o rádio está sintonizado, a informação de qual mão foi usada torna-se irrelevante. A nossa intuição é capturada quando temos uma estrutura da seguinte forma:²⁰

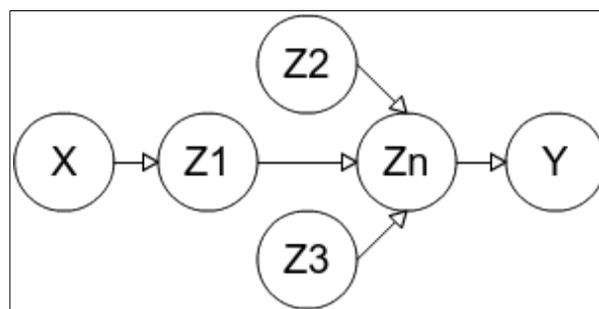


Figura 7: Problema da Transitividade

Pode ocorrer que existam meios diferentes (Z_1 , Z_2 e Z_3) de alterar uma variável (Z_n), e embora uma variável atual (Z_1 que foi causada por X) seja aquela que cause o efeito (Y), não

²⁰ Aqui é outro caso onde o uso de variáveis binárias atrapalha a explicitação de conteúdo causal. Poderíamos usar uma variável com os valores {mão esquerda, mão direita, nenhuma mão} ou duas variáveis cada uma com dois valores {mão esquerda, nenhuma mão} e {mão direita, nenhuma mão}. Como cada escolha representacional pode refletir um interesse explicativo diferente, parece vantajoso que não construamos nas próprias noções causais quantos valores as variáveis podem assumir..

consideramos a variável que começa a estrutura atual (X) relevante com respeito ao efeito (Y), pois o valor da primeira variável (X) se torna irrelevante dada a disponibilidade de rotas alternativas ($Z_2 \rightarrow Z_n \rightarrow Y$ e $Z_3 \rightarrow Z_n \rightarrow Y$) à rota original ($X \rightarrow Z_1 \rightarrow Z_n \rightarrow Y$).

O mesmo problema de transitividade que descrevemos ocorre quando temos que distinguir entre causas por omissão legítimas de ilegítimas. Causação por omissão é quando um efeito ocorre porque algo não ocorreu, quando dizemos por exemplo que a não administração do medicamento pelo médico causou a morte do paciente. Um promotor de justiça poderia se interessar no caso médico e querer saber se houve uma relação causal ou correlacional, pois no primeiro caso o médico poderia ter cometido um crime. Se a não administração do medicamento causou a morte do paciente, poderíamos dizer que a contrafactual (2.4.4) “se o médico tivesse administrado o medicamento, o paciente teria sobrevivido” é verdadeira. Mas, por que não estaríamos interessados em dizer que (2.4.5) “se S, que mora numa tribo indígena em Bornéu, não cursou medicina e nunca ouviu falar do paciente, administrasse o medicamento, o paciente teria sobrevivido” é verdadeira? Ou, mesmo que seja, porque não nos interessamos nela?

Existem dois motivos para o nosso desinteresse. O primeiro é que situações como (2.4.5) descrevem variações muito grandes do caso atual. A estrutura de (2.4.5) é muito diferente da descrita por (2.4.4) por mais que o resultado final (a sobrevivência do paciente) fosse o mesmo. Como usualmente nosso interesse está em cenários que não diferem muito do cenário atual, consideramos (2.4.4) adequada e (2.4.5) inadequada. Obviamente, um conjunto de variáveis que incluísse as situações descritas por (2.4.5) poderia chegar à conclusão que o paciente sobreviveria, mas esse conjunto é bem diferente do conjunto descrito por (2.4.4). O promotor pode apelar para (2.4.4) para incriminar o médico, mas não a (2.4.5) para incriminar S. Trivialmente, o médico é responsável pela morte do paciente, enquanto S não é. Uma abordagem intervencionista nos permite ver detalhadamente como isso ocorre, pois nos permite ver como precisaríamos intervir em alterar diversas variáveis para trazer um resultado, inclusive variáveis as quais o promotor não poderia usar para incriminar S por não serem legalmente previstas. O segundo motivo é que certas causações por omissão falham em fornecer detalhes sobre por que o resultado teve seu resultado particular e não outro. Tomemos a contrafactual “se um meteoro caísse em mim enquanto escrevo este parágrafo, ele não estaria escrito”. O problema dessa contrafactual é que ela não diferencia nada entre as diferentes formas em que eu poderia escrever este parágrafo. Pensando no rádio, não pensamos que L (o botão liga/desliga) seja suficiente para explicar a saída de som do alto-falante porque a alteração do valor de L é insensível a diversas variações em S .

Esses dois motivos nos levam a necessidade de tornar preciso o conjunto de variáveis em que uma estrutura está inclusa. O que precisaremos para a causação contribuinte é que ela seja caracterizada relativamente a um conjunto de variáveis V em uma estrutura particular. Dada as considerações de como lidamos com a transitividade, Woodward caracteriza formalmente a noção de causa contribuinte, nomeando-a M para “teoria manipulacionista”, do seguinte modo:

(M) Uma condição necessária e suficiente para X ser uma causa direta (em nível) de Y em respeito ao conjunto de variáveis V é que há uma possível intervenção em X que alterará Y , ou a distribuição probabilística de Y , quando mantemos fixos em algum valor todas outras variáveis Z_i em V . Uma condição necessária e suficiente para X ser uma causa contribuinte (em nível) de Y com respeito ao conjunto de variáveis V é que (i) há uma rota direcionada de X para Y tal que cada nó nesta rota é uma relação causal direta: isto é, um conjunto de variáveis $Z_1...Z_n$ tal que X é uma causa direta de Z_1 , que por sua vez é uma causa direta de Z_2 , que por sua vez é uma causa direta de... Z_n , que é uma causa direta de Y , e que (ii) há alguma intervenção em X que altera Y quando todas outras variáveis em V que não estão nesta rota são fixados em algum valor. Se há apenas uma rota P de X para Y ou se a única rota alternativa de X para Y além de P não contem variáveis intermediárias (i.e., diretas) então X é uma causa contribuinte de Y enquanto houver alguma intervenção em X que altere o valor de Y , para algumas das outras variáveis em V . (2003, p. 59)

Podemos dizer que a noção de causa total é menos refinada que a noção de causa contribuinte, no sentido em que a primeira incorpora diretamente **CS** e **CN**: “**(TC)** X é uma causa total de Y se e somente se há uma possível intervenção em X que alterará Y ou a probabilidade distributiva de Y .” (idem, p. 51) e com isso tem um menor escopo de aplicação, pois se aplicaria principalmente em modelos sem óbvias variáveis não expressas que romperiam a estrutura, tais como descritos por leis estritas ou por sistemas simplificados que não incluem muitas variáveis. Sistemas mais complexos normalmente exigem modelos mais detalhados, e para eles será adequado o uso da causalidade contribuinte para um adequado entendimento da estrutura causal, tanto que a formulação dos critérios para uma intervenção (seção 2.5) usará sempre a noção de causa contribuinte.

Com as noções causais definidas, **TC** e **M** são suficientemente expressivas para expressarmos as diferentes relações causais de tipo. Woodward deriva as noções causais de tokens (idem, pp. 74-86) e probabilísticas (idem, pp. 61-65) da sua abordagem de tipo, mas não as trataremos aqui. O restante deste capítulo consiste em especificações de assuntos relacionados à causação de tipo.

2.5 Intervenção caracterizada precisamente

Para determinamos uma intervenção entre duas variáveis, X e Y , que representem propriedades diferentes e incompatíveis, representadas por alguma unidade especificável, com o intento de determinar se X por algum processo altera Y , então I é uma variável de intervenção (**IV**) para X em respeito a Y se e somente se I atende as seguintes condições:

(IV)

I.1 I causa X .

I.2 I age como um alterador para todas outras variáveis que causam X . Isto é, certos valores de I são tais que, quando I atinge tais valores, X cessa de depender dos valores de outras variáveis que causam X e, passa a depender somente do valor atribuído por I .

I.3 Qualquer rota direcionada de I para Y passa por X . Isto é, I não causa diretamente Y e não é uma causa de qualquer das causas de Y que são distintas de X exceto, é claro, para aquelas causas de Y , se alguma, que estão construídas na própria conexão $I \rightarrow X \rightarrow Y$; isto é, exceto para a) qualquer causa de Y que são efeitos de X (i.e., variáveis que estão causalmente entre X e Y) e b) qualquer causa de Y que esteja entre I e X e não tem efeito em Y independentemente de X .

I.4 I é (estatisticamente) independente de qualquer variável Z que cause Y e que está em uma rota direcionada que não passa por X . (2003, p. 98)

Passemos às explicações de cada uma das condições. Trivialmente, **I.1** requer que I cause X . **I.2** requer que a intervenção seja caracterizada de maneira a isolar a relação causal da dependência de outras variáveis da estrutura em que ela está localizada. Quando expusemos informalmente a teoria (seção 2.2) mencionamos que precisávamos ajustar as variáveis do rádio (ele devia estar ligado, B em AM...) para verificarmos a relação $DI \rightarrow S$. Formalmente, precisamos que a intervenção seja caracterizada de maneira a isolar a relação da sua estrutura.²¹ A condição **I.3** exige que nossa intervenção não seja a causa do efeito. Queremos que a intervenção seja uma causação em corrente ($I \rightarrow X \rightarrow Y$) e evitar que a intervenção seja, na realidade, uma causa comum de dois efeitos ($X \leftarrow I \rightarrow Y$). Se quiséssemos saber se dedos amarelados causam câncer pulmonar, uma intervenção que fizesse as cobaias fumar seria mal desenhada porque obteríamos uma estrutura de causa comum, pois fumar estaria causando o amarelamento, mas o amarelamento não estaria causando o câncer pulmonar. Por último, a condição **I.4** exige que I não seja causada por Z que diretamente causa Y . Queremos evitar a seguinte estrutura “triangular”:

²¹ Pearl já introduzia o “operador *do*” [fazer] com essa finalidade. Na linguagem dele, o operador deveria ajustar a variável X em algum valor x , ou $do(X = x)$, onde os valores desse operador deveriam ser calculados para “mutilar” o gráfico onde X se encontrava (Pearl (2000) apud Walker & Gopnik (2013, pp. 347-348)).

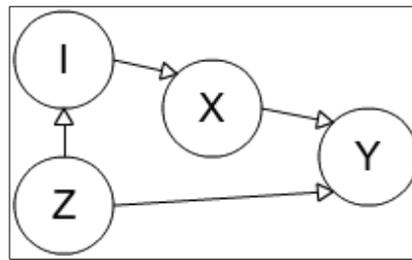


Figura 8: Causação Triangular

Este gráfico possui uma estrutura bastante próxima àquela da Fig. 6 (página 25) que mostrava a correlação entre a probabilidade de desenvolvimento de trombose, pilulas anticoncepcionais e gravidez. Como naquele caso, se não isolarmos a causa comum ($I \leftarrow Z \rightarrow Y$) dos seus efeitos do modelo, não poderemos determinar se a variável da intervenção é causalmente relevante para o efeito, i.e., se $I \rightarrow X \rightarrow Y$.

2.6 Reduccionismo

Fornecer um reduccionismo causal é fornecer uma tradução do conjunto de conceitos modais (causa, lei, contrafactualmente suportado...) para conceitos não modais. Quando oferecemos tal redução, dizemos algo específico sobre as variáveis às quais associamos causas e efeitos; falamos algo sobre o conteúdo, ou o tipo, das variáveis que estamos dispostos a atribuir causalidade; dizendo com isso que elas devem ser do tipo redutíveis ao que queremos reduzir. Como não vejo motivos prévios para limitar as variáveis que usamos, não vejo motivo para buscar um reduccionismo causal. Buscar um reduccionismo causal parece ser buscar algo que limita o escopo das relações as quais estaríamos predispostos a atribuir causalidade. Dizer que causalidade envolve transferência de energia espacotemporalmente rastreável é restringir os tipos de variáveis que poderiam caracterizar uma relação como causal. Assim como também seria dizer que a causalidade envolve disposições localizáveis, ou semelhança ostensiva aos casos onde nossas ações são causas. Como eu não veria problemas em aceitar uma relação causal se me mostrassem como as diversas causas estão funcionalmente relacionadas com as variáveis associadas ao efeito, não vejo razões para pensar que há algo específico, que se fosse reduzido, explicaria melhor a causalidade. Isto não é querer dizer que o reduccionismo causal seja um campo que nunca deva ser explorado, pois talvez até expliquem as condições suficientes para que uma situação específica, como em uma determinada área científica, seja uma relação causal. O problema que vejo é em mostrarem

como os critérios de redução que exigem para as variáveis sejam necessárias, pois eu sempre estaria disposto a recusar o que me oferecem como critérios se houvesse uma relação causal que atribuísse uma relação funcional adequada (que atendesse aos critérios da seção 2.4), mas que não possuísse a redutibilidade proposta. Os contraexemplos a abordagens reducionistas surgem justamente desta direção.

Por fim, parece vantajoso possuímos uma concepção deflacionada da causalidade porque ela poderá se adequar de maneira natural ao avanço de nossas teorias científicas. Um critério causal já defendido era o de contiguidade espacial: existiria uma continuidade espacial da causa ao efeito, de modo que causas não poderiam estar localizadas espacialmente distantes do seu efeito. Por tal critério, as leis newtonianas falhariam em explicar relações causais entre os astros porque suporia relações espacialmente não contíguas. Ficamos com um impasse: ou ficamos com a teoria que melhor explica os fenômenos que se propõe, ou a recusamos por não atender a um critério que caracterize relações causais.

Outro critério tradicionalmente utilizado para caracterizar a causação foi o de dependência temporal – as causas são temporalmente anteriores aos seus efeitos. Recentemente este critério suscitou um problema similar ao da contiguidade espacial, pois nossa atual teoria microfísica toma o tempo como sendo simétrico nas relações microfísicas: as leis se adequariam tanto se o fenômeno ocorresse do passado para o futuro quanto do futuro para o passado. O exemplo tradicionalmente usado é o da propagação da radiação, onde podemos descrever a expansão ondulatória esférica proveniente de uma fonte localizada pontualmente. O processo contrário, onde as ondas retornam à fonte obedeceria igualmente as nossas leis naturais. Por que preferir a expansão à contração? Ou a assimetria temporal no lugar da simetria?²²

Não queremos desconsiderar a assimetria temporal nas relações causais macrofísicas. Também não queremos sugerir que eventos macrofísicos são apenas eventos microfísicos que não sabemos como reduzir, mas que se a redução ocorresse haveria uma transferência da simetria temporal para os eventos macrofísicos. Essas últimas duas propostas estão fora dos nossos objetivos aqui. O nosso ponto é simplesmente que uma teoria deflacionada, que tenta descrever a causação apenas como relações entre variáveis, se mantém aberta a esses diferentes critérios e pode se adequar às nossas teorias científicas sem impô-los previamente. Da mesma maneira que ocorreu com a ação à distancia, podemos deixar a questão em aberto e manter certo conceito de causa.

Como damos conta desse impasse entre teorias científicas e conteúdo causal prévio? Se

22 Price (1992), pondo o problema em perspectiva, discute-o em maiores detalhes.

temos uma teoria que foca em descrever como variáveis têm de se comportar não precisamos ter enviesamento prévio se a relação causal é temporalmente assimétrica e espacialmente contígua – e deixamos a questão em aberto para nossa melhor ciência. A única coisa que precisamos ter clareza é sobre o que seria para as variáveis envolvidas na relação causal ter o seu valor alterado (o que discutiremos na seção 2.8), que é o que as teorias científicas buscam fazer. Em casos da vida comum pensamos na causalidade como temporalmente assimétrica porque as variáveis que utilizamos (supostamente) possuem tal assimetria, mas se possuímos uma teoria causal deflacionada, como a que apresentamos, não precisamos estender esse critério para fora da vida comum.

2.7 Invariância e Leis

Nosso objetivo nesta seção é mostrar como podemos usar uma noção mais fraca do que a noção de lei, a noção de Invariância. Invariância é uma propriedade robusta possuída pelas funções que associam os valores das variáveis de relações causais. Para diferentes funções, F_1 e F_2 , que descrevem uma relação causal $X \rightarrow Y$, uma função F_1 é mais invariante que uma função F_2 , se F_1 descrever: i) mais valores que possam ser assumidos por X e Y ; e/ou ii) menos valores fora da relação $X \rightarrow Y$ que romperiam a relação $X \rightarrow Y$. Usando um exemplo de Woodward (2003, p. 286), as Leis de Newton fornecem uma função menos invariante para explicarmos as relações causais entre os movimentos de objetos do que a Teoria da Relatividade Geral, porque as primeiras se romperiam,²³ i.e., deixariam de explicar o que ocorreria, se: os campos gravitacionais fossem fortes, onde um valor assumido por uma variável inclusa na lei romperia a explicação; e quando a velocidade (que é omitida pelas Leis de Newton) não é baixa em comparação com a velocidade da luz.

Estruturas explicativas respondem a um número maior ou menor de questões sobre o que teria acontecido se as coisas fossem diferentes, isto é, se alterássemos os valores das variáveis inclusas em estruturas explicativas. Quando uma estrutura explicativa é um subconjunto derivável de outra estrutura, dizemos que a última é mais explicativa que a primeira e é com isto que capturamos o grau de invariância.

Próximo ao ponto acima, quando uma estrutura explicativa incorporar outra, ao

²³ Vale notar que usamos “romper” e “isolar” de maneira diferente. Os ajustes nos botões do rádio para testar a relação $DI \rightarrow S$ a isolam da estrutura, mas não rompem a estrutura; enquanto alterar o país onde o rádio se encontra romperia a relação $DI \rightarrow S$, tal como descrita pelas associações que fizemos com estações de rádio. Com as Leis de Newton poderíamos calcular a trajetória de dois astros isoladamente da existências de outros astros, e neste caso a relação não é rompida.

responder questões sobre o que teria ocorrido nesses pontos de ruptura, ela será mais invariante. Para ilustrar esse ponto Woodward (2003 pp. 257-269) discute um exemplo fornecido por Haavelmo, no qual comparamos duas estruturas explicativas:

(2.8.1) A relação entre a velocidade do carro e a pressão exercida no acelerador em uma pista plana.

(2.8.2) As relações expressas por teorias de engenharia (incluindo diferentes princípios de funcionamento para mecanismos internos do carro e leis físicas e químicas) que explicam o funcionamento do carro, e decorrentemente a relação entre a pressão no acelerador e a velocidade do carro.

Haavelmo dizia que (2.8.1) possui menos autonomia que (2.8.2), e traduzindo aqui, poderíamos dizer que (2.8.1) é menos invariante que (2.8.2).²⁴ O ponto é que com (2.8.2) podemos explicar tudo que é explicado por (2.8.1), mas não o contrário; não podemos usar (2.8.1) para explicar (2.8.2).

Vale notar que não é apenas por podermos derivar um subconjunto de um conjunto maior, que este último conjunto é mais explicativo. Poderíamos ter teorias que explicam fenômenos diferentes conjuntas sem uma melhor explicação de por que elas funcionam. Embora uma conjunção da lei de Galileu com a lei da oferta e demanda permita derivarmos qualquer teoria da conjunção, a conjunção não explica melhor os fenômenos descritos individualmente pelas leis porque ela responderia ao mesmo número de questões sobre o que ocorreria se as coisas fossem diferentes. O mesmo não ocorre com a incorporação da lei de Galileu pelas leis de Newton que explica: tudo que é explicado pela primeira, o que ocorre quando uma variável explícita assume um valor não previsto (se o objeto em queda está muito longe da superfície terrestre), e o que ocorre quando a estrutura galileiana é rompida (por exemplo, se o objeto caísse em um planeta com massa diferente do nosso). A Lei de Galileu é derivável da newtoniana, mas o número de questões que as leis newtonianas respondem sobre o que ocorreria se as coisas fossem diferentes é maior.

Usando a invariância ganhamos uma maneira de quantificar a qualidade das explicações: explicações melhores respondem a um número maior de questões sobre o que teria acontecido se as coisas fossem diferentes. Passaremos a três exemplos. No primeiro interpretamos uma lei como sendo uma invariante²⁵, no segundo mostramos uma função invariante que é suportada por uma lei e por último mostramos uma função invariante da qual

24 Na passagem discutida em Woodward o conceito de “autonomia” não apresenta diferenças significativas do que chamamos de “invariância” aqui. O único ponto de desacordo é que Haavelmo parece sugerir que podemos pensar em um sistema com absoluta autonomia instrumentalmente, enquanto Woodward não pensa que seja necessário.

25 Para simplificação textual, também usaremos “invariante” como abreviação para “função invariante”.

desconhecemos leis de fundo.

Invariantes permitem incorporar generalizações descritas por leis. Tomemos a Lei de Ohm, onde a tensão (V) é igual à resistência (R) multiplicada pela corrente (I), ou seja $V=R*I$ ou $I=V/R$ ou $R=V/I$. A lei nos fornece as informações pelas quais a alteração de uma variável afeta as outras. X causar Y para a lei de Ohm é uma maneira de rastreamos como a alteração dos valores das variáveis (causa) acarretam um resultado (efeito). A alteração não precisa fazer referência à agência humana, pois as variáveis associadas independem de que um ser humano as altere. O que a invariante exige é que, se a alteração da variável associada a causa ocorresse, teríamos uma variação da causa associada ao efeito. Essa alteração obedece a um padrão, que, então, é invariável. No caso de leis estritas, como a lei da conservação de energia, temos um padrão que é invariável independente do contexto ou manipulação. De maneira bem linear, podemos usar $V=R*I$ como uma versão de $Y=F(X_1...X_n)$. Por exemplo, se mantivermos uma corrente constante de $1A$, podemos descrever a alteração da resistência ($R_1...R_n, X_1...X_n$) como a causa da alteração da tensão (Y, V): Se $R=1$, então $V=1$; se $R=33$, então $V=33$; se $R=470$, então $V=470$; e assim por diante.

Leis naturais formam um subconjunto, mesmo que muito especial e paradigmático, de generalizações capturadas por funções invariantes. Para nossa abordagem, não há dúvida de por que leis são importantes: elas são os casos de menor invariância possível sob manipulações. Uma generalização altamente variante permite pouca previsibilidade, e por contraste, uma generalização altamente invariante permite muita previsibilidade. Em razão desta gradação de invariância existirá um gradiente entre o poder de previsão das generalizações que são suportadas por leis (em um extremo do gradiente) e diversas generalizações menos preditivas à medida em que se afastam desse extremo. Dados os objetivos explicativos podemos escolher quanta previsibilidade é necessária e reforçar ou enfraquecer a explicação que usamos. O que achamos importante é a eliminação de uma possível dicotomia entre: generalizações suportadas por leis, onde existiria previsibilidade ou necessidade; e generalizações não suportadas por leis, onde inexisteria previsão e haveria mera especulação. A aplicação com os próximos dois exemplos ilustrará esse gradiente.

As relações descritas pelo nosso exemplo do rádio são, em grande parte, suporta pela lei de Ohm. Quando giramos $D3$ em sentido horário, diminuimos a resistência do dial, que sendo um potenciômetro é um resistor variável, e com uma menor resistência nele, a potência em S aumenta, aumentando o volume do rádio. Mas, como pretendemos ter sugerido, o modelo do rádio expresso pelas equações estruturais da seção 2.2 já é suficiente para fornecer explicações sobre o funcionamento do rádio. Além disso, tem certos detalhes do que podemos

extrair do rádio que não são adequadamente capturados em uma linguagem estritamente nômica, como a relação entre as posições de DI e as estações específicas: não há uma lei natural estrita que especifique os anos nos quais a UFRGS AM estará no ar.

Passando ao último ponto, podemos ter generalizações invariantes sob manipulações sem qualquer conhecimento de leis subjacentes às generalizações, e tais generalizações ainda fornecem informação para contrafactuais. Um mercador pode alterar a disposição dos produtos nas gôndolas de seu estabelecimento e rastrear o que acontece em cada disposição. Se descobrir como as disposições dos produtos afetam as vendas, pode usar essa informação para aumentar suas vendas sem precisar rastrear qualquer lei que suporte a informação. É de se esperar que pesquisadores de marketing busquem generalizações (algo como leis) entre disposição de produtos e vendas. Mas uma generalização invariante sob manipulações já é suficiente para ajudar tanto o pesquisador quanto o mercador, e nisto queremos mostrar outra vantagem de usarmos invariantes. Um pesquisador poderia buscar quais são as melhores disposições e fornecer informação de caráter genérico para diversos estabelecimentos. Apesar disso, a informação fornecida pelo pesquisador pode não funcionar para o mercador. O mercador pode então verificar quais são as alterações próprias que em seu estabelecimento aumentam as vendas. Diríamos que enquanto o pesquisador busca informações para um grande número de estabelecimentos, o mercador somente procura para o seu. A noção de invariância pode se aplicar em nível (*type* ou *level*), como no caso do pesquisador, ou em particular (*single* ou *token*), como no caso do mercador. O pesquisador e o mercador possuem informação para suporte de contrafactuais que é derivada das generalizações corroboradas pelos seus testes, mas não necessitamos de lei alguma. Além disso, a informação particular do mercador, mesmo não possuindo a generalidade da informação do pesquisador, é a que lhe importa. Tanto para o mercador quanto para o pesquisador, alterar a disposição dos produtos (causa, X) é uma maneira de alterar as vendas (efeito, Y). Evidentemente esta relação não possui estabilidade irrestrita sob manipulações, e não é fácil determinar precisamente as situações que seria rompida ou alterada, como a mudança do perfil do consumidor do mercado ou a situação econômica da região.

A noção de invariante permite trabalhar diretamente com esses casos, e então fornecer informação de suporte para contrafactuais sem precisar recorrer a um melhor sistema de leis naturais como proposto por Lewis. Do mesmo modo que alguém que desconhecesse o funcionamento do rádio da seção 2.3 poderia determinar seu funcionamento intervindo nas variáveis (botões) que achasse capaz de alterar S , um mercador que desconhecesse os detalhes da relação entre disposição de produtos e maximização de vendas poderia trabalhar com seu

caso concreto para descobrir a estrutura adequada para o seu mercado.

2.8 Alterando o Valor de Uma Variável

Como já mencionamos, a teoria intervencionista não pretenderá ser reducionista sobre a causalidade, i.e., ela não pretenderá traduzir termos modais em termos não modais. Defendemos aqui que as relações causais são relações funcionais entre variáveis que associam causas a efeitos de maneira invariante. As alterações associativas nos valores das variáveis são aquelas que podemos descrever como intervenções segundo os critérios **I.1-I.4**. E, podemos conhecer o conteúdo de certas intervenções, i.e., saber o que ocorreria se as coisas houvessem sido diferentes, conhecendo o conteúdo de outras intervenções. Todas as noções modais que usamos (relação funcional invariante entre variáveis, causa direta, causa total, causa contribuinte...) não possuem uma tradução em termos não modais, o que precisamos ter clareza é sobre *o que é para uma variável ter seu valor alterado*, e é sobre isso que nos ocuparemos nesta seção.

Peguemos o famoso exemplo de Quine (2013, 1960, p. 203) que criticava a semântica por trás de contrafactuais, e conseqüentemente teorias que as usassem. Isso ocorreria por que não conseguiríamos dar uma interpretação para contrafactuais como:

(2.9.1) Se Júlio César fosse o encarregado do exército estadunidense na guerra da Coreia, ele teria usado a) armas nucleares; ou b) catapultas.

Para a teoria intervencionista não damos conta de todas as contrafactuais. As únicas contrafactuais que temos como decidir o valor de verdade são aquelas para as quais temos uma ideia clara do que seria para as variáveis utilizadas trocarmos de valor. Contrafactuais como (2.9.1a) e (2.9.1b) envolvem alterações das quais não temos uma ideia clara do que seria ter seu valor alterado: não sabemos o que seria uma intervenção que pudesse determinar o que seria alterar alguém para ser Júlio César.

Para responder a crítica de Quine mais precisamente falaremos do problema que constantes colocam quando usadas em explicações, e posteriormente responderemos diretamente à sua crítica. Decorre da abordagem intervencionista que constantes não são suficientes para explicar os valores de verdade de contrafactuais. Sempre que uma alegação envolver uma constante precisamos recorrer a uma variável para explicá-la. A ideia parece trivial quando pensamos que é insuficiente dizer (2.9.2) “o mantimento do efetivo policial causou o aumento da criminalidade” para explicar o aumento da criminalidade. Só pensamos

que (2.9.2) é explicativa quando possuímos uma outra relação entre variáveis que tenha se alterado, como por exemplo o aumento do tráfico de drogas ou o aumento da desigualdade social da população. Também temos contrafactuais mais difíceis que (2.9.2) por possuírem várias interpretações. Woodward (2003, pp. 114-117) discute o seguinte exemplo:

(2.9.3) Ser mulher causa discriminação em respeito ao salário e/ou contratação.

Naturalmente, pensamos que alegações causais acarretam uma contrafactual “não-não”, de forma que para (2.9.3) teríamos “se S não fosse uma mulher, não haveria discriminação em respeito ao seu salário e/ou contratação”. O problema aqui é que, como usamos uma constante, tornamos obscuro qual seria a contrafactual não-não correta por trás de (2.9.3). Embora existam meios de alterar o sexo de pessoas, a contrafactual (2.9.3) normalmente não é tomada em sentido literal. Podemos, apesar dessa dificuldade, esclarecer o conteúdo de (2.9.3) através das possíveis associações de (2.9.3) com diferentes intervenções que poderiam lhe testar. Comparemos três possíveis contrafactuais conectadas a alegação causal (2.9.3):

(2.9.3a) Se S se tornasse um homem, não haveria discriminação em respeito ao seu salário e/ou contratação.

(2.9.3b) Se as crenças dos empregadores fossem diferentes, o salário e/ou contratação de S não seria afetado por seu sexo.

(2.9.3c) Se nosso sistema econômico não fosse discriminatório, S receberia o salário ou seria contratada de maneira igualitária aos homens do sistema.

Usa-se habitualmente (2.9.3) para tentar mostrar como a variável “ser mulher” é causalmente determinante quando mantemos outras variáveis constantes (formação, experiência profissional, idade). Não interpretarmos (2.9.3) como significando (2.9.3a) porque não é essa contrafactual que nos interessa; não queremos saber se cirurgias para alterações de sexo fossem realizadas, pessoas que deixassem de ser mulheres passariam a receber maiores salários. A interpretação mais corrente é (2.9.3b), que nos diz que o problema está no enviesamento dos empregadores. Podemos pensar em testes com conteúdos bem definidos para (2.9.3b), como fornecer aos empregadores currículos distribuídos em três grupos: a) com os sexos mantidos; b) sem informação sobre o sexo; c) com os sexos invertidos. Se (2.9.3b) for verdadeira podemos esperar que mulheres recebam menos do que homens no grupo a), igualmente no grupo b), e mais e no grupo c). A última interpretação, (2.9.3c), diz respeito a um sistema de distribuição de renda não igualitário sendo comparado com um igualitário. Como o que estaria em jogo seria a relação dos indivíduos a diferentes sistemas econômicos precisaríamos conhecer sistemas igualitários que fossem adequados para fazermos o contraste. O problema de (2.9.3c) é que temos pouco conhecimento de sistemas econômicos que difiram

somente com respeito à discriminação. Talvez (2.9.3c) possa ser esclarecida através de algum teste de intervenção, assim como (2.9.3b), mas ela sozinha torna difícil de tornarmos mais clara a alegação (2.9.3).

Muitos tomariam (2.9.3) como verdadeira por simplesmente descrever um fato corrente, mas determinar se a alegação é causal, e seu valor de verdade, envolve conseguir descrever o que ocorreria se alterássemos as próprias variáveis da alegação e não outras. Envolve a possibilidade de construirmos um modelo explicativo que nos informe sobre como diversas variáveis são relevantes a outras variáveis. O problema não é o uso de (2.9.3), ou que exista quem diga verdadeiramente que “ser uma zebra causa surgimento de listras”, o ponto é que deveríamos preferir contrafactuais mais claras, e as mais claras são a que nos informam o que ocorreria se a própria variável da contrafactual alterasse seu valor dentro de um sistema que contenha outras variáveis que sejam relevantes às alegações.

Essa crítica pode ser considerada como dizendo respeito a formulações das contrafactuais, e que, dado o interesse de quem pronuncia (2.9.3), podemos entender quais das três contrafactuais que derivamos seria a pretendida. Mas o ponto pode ser ampliado. Quando escolhemos (2.9.3b), que normalmente é a escolhida, e recusamos (2.9.3a), recusamos que ser mulher esteja *causando* a discriminação, e aceitamos que a discriminação dos empregadores está *causando* pagamentos e contratações discriminatórios.

Escolher a contrafactual adequada, e saber o que está em jogo para teste, é possibilitar ir em direções que diminuam discriminação. Assim como as diferentes relações causais ilustradas pela Fig. 2 ($Y \leftarrow X \rightarrow Z$, $X \rightarrow Y \rightarrow Z$ e $X \rightarrow Z \leftarrow Y$) informavam que diferentes intervenções trariam resultados diferentes, diferentes construções da relação ternária entre o sistema econômico discriminatório, o enviesamento dos empregadores e a discriminação em respeito ao salário das mulheres, acarretarão resultados diferentes sob intervenções.

O problema da alteração das variáveis também aparece para abordagens tradicionais de leis naturais que defendem que leis devem expressar condicionais universalmente quantificados. Para uma lei “Todos *As* são *Bs*” deveríamos expressar uma contrafactual do tipo “se um *X* que não é um *A*, fosse um *A*, então seria um *B*”. E novamente parece que temos de dar conta de contrafactuais estranhas. Num exemplo inspirado em Morton (1973, p. 323), se tivéssemos a lei “Todo fósforo seco quando riscado acende” teríamos que dar conta de contrafactuais como “Se eu fosse um fósforo seco, acenderia quando riscado”. Mas como viemos sugerindo, o problema não é que contrafactuais sejam obscuras, o problema é que não temos como determinar o valor de contrafactuais que dizem respeito a alterações das quais não temos conhecimento. Contrafactuais para as quais temos uma ideia clara do que é para

suas variáveis terem seus valores alterados (como girar botões de rádio) possuem um conteúdo muito claro.

Adquirimos uma ideia clara do que seria para uma variável ter seu valor alterado tanto na nossa vida comum quanto da ciência. O repertório causal básico de crianças lhe ensina que relações causais se dão, normalmente, espacialmente próximas, tal que esperam que uma bola se mova porque outra lhe bateu. Quando aprendem a usar um controle remoto aprendem que a relação que elas podem intervir não precisa de contato espacial e com isso reajustam seu mapa causal sobre o mundo. Adultos interrogados sobre acidentes aéreos não consideram contrafactuais sobre a gravidade espontaneamente cessando de operar, e mesmo nas contrafactuais que consideram, teriam dificuldade de expressar contrafactuais em que o acidente aéreo pudesse ser evitado por desconhecimento sobre o assunto (WALKER & GOPNIK, 2013, p. 353). Esses exemplos sugerem que o problema não está na nossa noção de causa ou no pensamento contrafactual que usamos a todo momento; o problema está em contrafactuais das quais não temos conhecimento do que seria para as variáveis descritas terem seus valores alterados. Se, assim como as crianças que aprendem a usar controles remotos, aprendêssemos adequadamente o que está por trás de contrafactuais obscuras como (2.9.1), poderíamos determinar seu valor de verdade. Enquanto não sabemos o que é para alguém vir a ser Júlio César, não podemos nos comprometer nem com (2.9.1a) nem com (2.9.1b).

CAPÍTULO 3 – ENFRENTANDO AS CRÍTICAS FEITAS AO MODELO ND

3.1 Introdução

Todas as caracterizações feitas no segundo capítulo (CN e CS para TC, CD, NC* e M, e I.1-I.4) caracterizam o que é para uma relação ser causal e são normativas para a explicação causal: uma explicação causal que não atenda tais critérios não será satisfatória. Quando lidamos com os contraexemplos internos do modelo ND vemos que eles não atendem aos critérios de caracterização de uma relação causal. Quanto às críticas externas, vemos que a noção de Invariância captura o papel normalmente ocupado pelas Leis Naturais, podendo ser aplicada de modo direto.

Nos ocuparemos com explicações causais, que são aquelas em que relacionamos alterações de variáveis – onde pela alteração do valor da causa há alteração do valor do efeito. Não nos ocuparemos com explicações puramente descritivas, como “corvos são negros”, ou propostas por ciências formais, como Matemática e Lógica.²⁶ A justificativa para isso é que as explicações para as quais o modelo ND pretendia se aplicar eram aquelas em que fosse possível rastrear como alterações no *explanans* eram relevantes para alterações no *explanandum*.

Para a abordagem intervencionista da causação toda relação causal é contrastiva, e não existem relações causais sem a possibilidade de contraste.²⁷ Uma generalização como “todos corvos são negros” não é explicativa por não dizer nada sobre o que faz corvos serem negros e não nos informa por que deveríamos esperar que fossem. Por outro lado, as relações químicas que pigmentam a plumagem, as diferentes associações entre outras aves e suas cores derivadas de pressões seletivas, o papel que a cor pode ter para o ambiente ou para escolha de parceiro sexual são informações invariantes sob manipulações e que descrevem relações causais; são essas informações que respondem a questões sobre o que teria acontecido se as coisas fossem diferentes e é por isso que as consideramos explicativas. A generalização “todos corvos são negros” talvez seja derivável dessas informações, mas a generalização não é explicativa por si.

26 Embora essas informações façam parte de explicações causais (por exemplo, usamos matemática na lei de Ohm) elas não são o que há de causal em tais explicações.

27 Como Schaffer (2005) aponta, muitos problemas com relações causais são evitados quando formulamo-las em alegações contrastivas.

3.2 Críticas ao modelo ND

3.2.1 Assimetria Explicativa

A assimetria explicativa ocorria nos casos em que se trocássemos o *explanans* de lugar com o *explanandum* deixávamos de ter uma explicação adequada para o *explanandum*. Naquele caso, poderíamos calcular o tamanho de um mastro usando sua sombra ou o tamanho da sombra usando o tamanho do mastro. Ampliando o exemplo, para a abordagem intervencionista não é qualquer relação derivável de um sistema formal que se caracteriza como uma relação causal – como em qualquer cálculo onde se desconhece o valor de uma variável e pode-se descobrir outra.

O problema do mastro é que enquanto existem alterações no tamanho do mastro que alterariam o tamanho da sombra, isso não ocorre para o mastro. A única alteração possível do tamanho da sombra seria através da alteração do mastro, mas se fizéssemos essa alteração violaríamos a condição **I.3**.

3.2.2 Irrelevância Explanatória

A irrelevância explicativa ocorria nos casos onde atendíamos aos critérios do modelo ND, mas a informação do *explanandum* não era relevante. Naquele caso tínhamos a lei “todos homens que tomam pílulas anticoncepcionais não conseguem engravidar”; a condição inicial que instanciava a lei “Jones é um homem que toma pílulas anticoncepcionais”; e o *explanandum* que se seguia dedutivamente “Jones não consegue engravidar”. O problema é que não atendíamos ao critério mínimo para caracterizarmos uma relação causal onde precisaríamos que duas variáveis pudessem assumir dois valores diferentes, e para cada valor assumido houvesse uma alteração da causa para o efeito, o que colocamos formalmente com $y_1 = F(x_1) \neq y_2 = F(x_2)$.²⁸ E o caso de Jones é justamente um onde temos uma alteração de x_1 para x_2 sem uma alteração correspondente de y_1 para y_2 . Se repararmos, tanto na lei quanto no *explanandum*, não temos a caracterização de uma relação causal, e portanto não possuiremos uma explicação causal.

²⁸ Essa caracterização mínima está incorporada nas definições de todas noções causais – CN, CS, TC, CD, NC* e M.

3.2.3 Alegações Causais Singulares

Para a abordagem intervencionista não há problemas em caracterizar alegações singulares causais como explicativas quando caracterizadas causalmente, e, para fazermos a caracterização causal não precisamos do conceito de lei – que não foi usado em nenhuma definição causal. O contraste entre (1.1), que descrevia a relação causal entre a joelhada e a queda do tinteiro, e alguma explicação que incluía leis sobre comportamento de fluídos e forças necessárias para a queda do tinteiro, é próximo ao contraste entre (2.8.1), a relação da pressão no acelerador e aceleração do carro, e (2.8.2), relacionando diversos princípios de engenharia, leis físicas e informações geográficas. Alegações singulares, tanto (1.1) e (2.8.1), expressam relações invariantes sob manipulações que podem ser usadas preditivamente respondendo questões sobre o que teria acontecido se as coisas fossem diferentes; além disso elas são subconjuntos de explicações de maior invariância que as explicariam melhor. Mas, como já argumentamos na seção 2.8, a diferença entre (2.8.1) e (2.8.2) é de grau e não de tipo.

3.2.4 Leis e Ciências Especiais

As leis são explicadas pela abordagem intervencionista da causação por serem as relações mais exploráveis para intervenções: elas descrevem de maneira detalhada o que ocorreria se intervíssemos para alterar as variáveis da relação descrita pela lei. Ficamos com um contínuo de invariância: em um extremo temos leis paradigmáticas como a da conservação de energia e no outro alegações singulares como (1.1), mas no meio ainda temos diversas relações exploráveis para intervenções como a lei da oferta e demanda que expressa uma relação explorável de maneira mais detalhada e precisa que (1.1), mas menos do que a lei da conservação de energia. Embora as ciências especiais não possuam leis paradigmáticas, elas possuem relações invariantes sob manipulações, e é o caráter invariante que permite a previsibilidade necessária para explicações.

A diferença entre as explicações fornecidas pelas ciências especiais das estritas, a Física e a Química, é a diferença entre a ruptura dos seus modelos. Existem mais variáveis não descritas que romperiam uma explicação feita pela lei da oferta e demanda do que pelas leis newtonianas. Contudo, a possibilidade de ruptura de uma explicação não torna o que já é explicado menos explicativo. Saber que interferência estatal no preço da gasolina romperia a relação de oferta e demanda não explica os casos onde não há interferência estatal; assim

como saber que as leis newtonianas não explicam a precessão anômala do periélio de Mercúrio não torna melhor a explicação do periélio terrestre.

CONCLUSÃO

Podemos com uma abordagem intervencionista dar conta dos contraexemplos tradicionais ao modelo ND. Os problemas surgiam pelo modo como os proponentes empiristas do modelo viam noções modais. Ficamos com uma análise de conceitos modais feita através de conceitos modais, e portanto circular. Consideramos apesar disso que o círculo de conceitos modais operativos da abordagem intervencionista da causalção é esclarecedor e temos certas dúvidas sobre os sucessos de uma possível redução. Nossa abordagem causal é deflacionada e deixa espaço para o nosso conhecimento científico determinar o que é para uma variável ter seu valor alterado. Embora esses sejam pontos disputáveis, pretendemos ter mostrado motivos favoráveis ao nosso lado da discussão. Ao possuímos uma teoria causal, podemos mostrar como os critérios para caracterização de relações causais podem ser usados para nos informar sobre o caráter explicativo de alegações causais. E como construímos os critérios usando variáveis, ficamos com critérios não misteriosos para serem usados.

Quanto às respostas aos contraexemplos, não é à toa que Hempel discutia seriamente com Scriven, e levava a sério as críticas sobre alegações que pareciam triviais, como joelhadas derrubando tinteiros; isto porque a linha limítrofe entre explicações científicas e não científicas não é algo fácil de se determinar. Não parece trivial dizer que a ciência não deveria se ocupar de explicar por que objetos caem. Também é difícil determinar o exato momento em que uma explicação passa a ser científica. Começamos de maneira quase inocente dizendo que existe uma relação entre a pressão exercida no acelerador e a velocidade do carro em uma linha reta; se formos adicionando informações mais detalhadas como aquelas da octanagem do combustível, o princípio de operação do motor à combustão, informações sobre a relação entre os pneus e o atrito de diferentes pistas... em que momento diremos que passamos de uma explicação não-científica para uma científica?

Como reconstruímos explicações apelando a invariantes sob manipulações e não apelando a leis, a explicabilidade de explicações (a quantidade de questões sobre o que teria acontecido se as coisas fossem diferentes que uma explicação fornece) será vista de maneira gradativa, e não temos como determinar o exato ponto em que uma explicação passa a ser científica. Isto pode ser considerado um ônus por aqueles proponentes do modelo ND que consideravam que o modelo fornecia os critérios demarcatórios para a cientificidade de uma explicação. Mas, dado os contraexemplos persistentes ao modelo ND e a maneira natural de reconstruímos explicações e explicar seu grau de previsibilidade, consideramos um bônus que a abordagem intervencionista da causalção não demarque rigidamente a linha entre ciência e

não-ciência.

Apesar das possíveis disputas que possam surgir sobre os pontos que acabamos de levantar. Vale notar que a teoria intervencionista é capaz de incorporar as vantagens que eram fornecidas pelo modelo ND. Primeiro, fornecemos critérios objetivos para determinar se uma explicação é boa. Segundo, fornecemos um tratamento unificado para todas explicações causais. E terceiro, fornecemos um modelo que é adequado à ciência, e dado o deflacionamento de certos critérios, adequável às alterações na ciência.

REFERÊNCIAS BIBLIOGRÁFICAS

- BEEBEE, HITCHCOCK & MENZIES. (Eds.) *The Oxford Handbook of Causation*. New York: Oxford University Press, 2009.
- CALL, J., & TOMASELLO, M. *Primate cognition*. New York: Oxford University Press, 1997.
- GOPNIK, A., GLYMOUR, C., SOBEL, D. M., DANKS, D., SCHULZ, L. E., & KUSHNIR, T. "A Theory of Causal Learning in Children: Causal Maps and Bayes Nets." Em: *Psychological Review*. Vol. 111, No. 1, pp. 3-32, 2004.
- HEMPEL, C. "The Function of General Laws in History" Em: *The Journal of Philosophy*, Vol. 39, No. 2, pp. 35-48, 1942.
- HEMPEL, C. *Filosofia da Ciência Natural*. Rio de Janeiro: Zahar Editores, 1974. Tradução: Plínio Sussekind Rocha.
- HITCHCOCK, C. "Causal Modelling". Em: *The Oxford Handbook of Causation*. New York: Oxford University Press, 2009.
- LEWIS, D. *Philosophical Papers II*. New York, Oxford: Oxford University Press, 1986. Traduções nossas.
- _____. *Humean Supervenience Debugged*. Oxford University Press, 1994. *Mind*, New.
- MORTON, A. "If I Were a Dry Well-Made Match" Em: *Dialogue* 12: 322-24, 1973
- MUNFORD, S. & ANJUM, R. L. *Causation – A very Short Introduction*. New York: Oxford University Press, 2013.
- _____. "Counterfactual Theories of Causation", Em: *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), Edward N. Zalta (ed.), Disponível em: <http://plato.stanford.edu/archives/spr2014/entries/causation-counterfactual/>. Último acesso: 20 de Maio de 2015
- MENZIES, P & PRICE. 'Causation as a Secondary Quality'. Em: *British Journal for the Philosophy of Science* 44: 187–203. Traduções nossas. Versão utilizada disponível em: http://sydney.edu.au/time/price/preprints/secondary_causes.pdf. Último acesso: 11 de Abril de 2015.
- PEARL, J. *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press, 2000.
- PRICE, H. "Agency and Causal Asymmetry". Em: *Mind*, New Series, Vol. 101, No. 403 (Jul., 1992), pp. 501-520. Traduções Nossas.

- QUINE, W. V. *Word And Object*. Massachusetts: MIT Press, 2013.
- RUSSELL, B. "On the Notion of Cause". Em: *Proceedings of the Aristotelian Society*, 13:1, 1917.
- SALMON, W. "Statistical Explanation." Em: *Statistical Explanation and Statistical Relevance*. Wesley Salmon (Ed.) Pittsburgh: University of Pittsburgh Press, 1971.
- SCHAFFER, J. "Trumping Preemption". Em: *The Journal of Philosophy*, Vol. XCVII, No. 4, (April 2000)
- _____. "Contrastive Causation". Em: *The Philosophical Review*, Vol. 114, No. 3 (July 2005)
- SCRIVEN, M. "Truisms as the Grounds of Historical Explanations." Em: *The Nature of Historical Explanation*, 443-75, Patrick Gardiner (Ed.). Oxford: Oxford University Press, 1959.
- SPIRITES, P., GLYMOUR, C. N., & SCHEINES, R., *Causation, prediction, and search*. New York: Springer-Verlag, 1993.
- WALKER, C. & GOPNIK, A. "Causality and Imagination" Em: *The Oxford Handbook of the Development of Imagination*. Marjorie Taylor (Ed.) New York: Oxford University Press, 2013.
- WOODWARD, J. - "Law and explanation in biology: Invariance is the kind of stability that matters". Em: *Philosophy of Science*, 68, pp. 1-20., 2001.
- _____. *Making Things Happen: A Theory of Causal Explanation*. New York: Oxford University Press, 2003. Traduções nossas.
- _____. "Interventionist Theories of Causation in Psychological Perspective". Em: *Causal Learning: Psychology, Philosophy and Computation*. GOPNIK, A. & SCULZ, L. (Orgs.). New York: Oxford University Press, 2007.
- WOODWARD, J. & HITCHCOCK C. - "Explanatory Generalizations, Part I: A Counterfactual Account*" Em: *Noûs* 37:1, pp. 1-24, 2003.

APÊNDICE A – Tabela relacionando diferentes posições do dial D1 à diferentes estações de rádio:

Frequência	Estação	Posição D1
540 kHz	Rádio Real AM (Canoas)	0,000
600 kHz	Rádio Gaúcha (Porto Alegre)	12,706
640 kHz	Rádio Bandeirantes (Porto Alegre)	21,176
680 kHz	Rádio Farroupilha AM (Porto Alegre)	29,647
720 kHz	Rádio Guaíba (Porto Alegre)	38,118
780 kHz	Rádio Caiçara AM (Porto Alegre)	50,824
900 kHz	Rádio ABC (Novo Hamburgo)	76,235
970 kHz	Rádio Pampa AM (Porto Alegre)	91,059
1020 kHz	Rádio Grenal AM (Porto Alegre)	101,647
1080 kHz	Rádio UFRGS AM (Porto Alegre)	114,353
1120 kHz	Rádio Rural AM (Porto Alegre)	122,824
1340 kHz	CBN (Porto Alegre)	169,412
1390 kHz	Rádio Esperança AM (Porto Alegre)	180,000

Fonte: https://pt.wikipedia.org/wiki/Lista_de_r%C3%A1dios_do_Rio_Grande_do_Sul.

Último acesso: 10 de Outubro de 2015