

## CAPÍTULO 3

# AVALIANDO RELAÇÕES CAUSAIS

### RESUMO:

A ciência política moderna é constituída fundamentalmente ao redor da tentativa de estabelecer relações causais entre conceitos importantes. Esse processo raramente é simples e serve como base para quase todas as controvérsias científicas. Como sabemos, por exemplo, se o desenvolvimento econômico causa a democratização ou se a democratização causa o desenvolvimento econômico, ou os dois, ou nenhum? Falando mais diretamente, se desejamos avaliar se algum  $X$  causa (ou não) algum  $Y$ , precisamos superar quatro barreiras: (1) Existe algum mecanismo causal crível que conecta  $X$  a  $Y$ ? (2) Podemos eliminar a possibilidade de  $Y$  causar  $X$ ? (3) Existe alguma covariação entre  $X$  e  $Y$ ? (4) Controlamos por todas as variáveis  $Z$  possíveis que podem causar uma relação espúria entre  $X$  e  $Y$ ? Muitas pessoas, especialmente as da mídia, cometem erros ao achar que ultrapassar a terceira barreira – observar que  $X$  e  $Y$  covariam – é a mesma coisa que conseguir ultrapassar as quatro. Em suma, encontrar uma relação não é a mesma coisa que encontrar uma relação causal, e é na causalidade que estamos interessados como cientistas políticos.

*Eu preferiria descobrir uma lei causal a ser o Rei da Pérsia.*

Demócrito (citado por Pearl, 2000)

### 3.1 A CAUSALIDADE E A LINGUAGEM DO DIA A DIA

Como a maioria das ciências, a disciplina da ciência política é fundamentalmente constituída em torno da avaliação de afirmações causais. Nossas teorias – que podem estar certas ou erradas – tipicamente especificam que alguma variável independente causa alguma variável dependente. Então nos esforçamos para achar evidências em-

píricas apropriadas para avaliar o quanto uma teoria é ou não suportada. Mas como avaliamos afirmações causais? Neste e no próximo capítulo, discutimos alguns princípios que podemos utilizar para fazer isso. Focamos na lógica da causalidade e em uma série de critérios para estabelecer com alguma confiança em que grau uma conexão causal entre duas variáveis existe. Então, no capítulo 4, discutiremos os vários tipos de desenho de pesquisa que nos ajudam a investigar afirmações causais. Na medida em que buscamos respostas para perguntas sobre relações causais, devemos ter em mente as “regras para o conhecimento científico” estabelecidas no capítulo 1, em particular a advertência sobre considerar apenas evidências empíricas no caminho.

É importante reconhecer a distinção entre a natureza da maioria das teorias científicas e o modo como o mundo parece ser ordenado. A maioria de nossas teorias são descrições limitadas a descrições de relações entre uma única causa (a variável independente) e um único efeito (a variável dependente). Tais teorias, nesse sentido, são representações simplistas da realidade, mas devem ser necessariamente assim. De fato, como notamos no final do capítulo 1, teorias desse tipo são louváveis em um aspecto: elas são parcimoniosas, tais quais pedaços de informação digeríveis. Não podemos enfatizar fortemente o bastante que a maioria das nossas teorias sobre fenômenos sociais e políticos é **bivariada** – isto é, envolve apenas duas variáveis.

Mas a realidade social *não* é bivariada; ela é **multivariada**, no sentido de que qualquer variável dependente de interesse é causada por mais de um fator. (“Multivariada” significa simplesmente “muitas variáveis”, o que quer dizer que envolve mais de duas variáveis.) Então, embora nossas teorias descrevam a relação proposta entre alguma causa e algum efeito, devemos ter sempre em nossas mentes que o fenômeno que estamos tentando explicar certamente possui outras possíveis causas. E, quando for o momento de desenharmos a pesquisa a fim de testar nossas ideias teóricas – tópico trabalhado no capítulo 4 –, temos que tentar explicar, ou “controlar para” essas outras causas. Caso contrário, nossas inferências causais sobre nossa teoria – de que X causa Y – podem muito bem estar erradas<sup>1</sup>. Neste capítulo, debruçamo-nos sobre alguns princípios práticos para avaliar se, de fato, algum X causa (ou não) Y. Estes critérios também são úteis para avaliar afirmações causais feitas por outros – sejam eles um jornalista, um candidato, um cientista político, um colega de classe, um amigo, ou qualquer outra pessoa.

Quase todo mundo, quase todo dia, usa a linguagem da causalidade – algumas vezes formalmente, mas na maior parte das vezes de maneira informal. Sempre que falamos de como um evento muda o curso de eventos subsequentes, invocamos o raciocínio causal. Mesmo a palavra “porque” implica que um processo causal está em operação<sup>2</sup>. Contudo, apesar da utilização generalizada das palavras “porque”, “afeta”, “impacta”, “causa” e “cau-

<sup>1</sup> Ao longo deste livro, tanto no texto como nas figuras, utilizaremos setas como sinal para “causalidade”. Por exemplo, o texto “ $X \rightarrow Y$ ” deve ser lido como “X causa Y”. Frequentemente, especialmente nas figuras, essas setas terão pontos de interrogação em cima delas, indicando que a existência de uma conexão causal entre os conceitos é incerta.

<sup>2</sup> Esse uso dos termos foi trazido à nossa atenção por Brady (2002).

salidade”, o significado dessas palavras não é exatamente claro. Filósofos da ciência têm realizado longos e vigorosos debates sobre formulações rivais de “causalidade”<sup>3</sup>.

Embora nosso objetivo não seja nos aprofundarmos tanto nesse debate, existe um traço dessa discussão sobre causalidade que merece uma breve menção. A maioria dos debates sobre filosofia da ciência se origina do mundo das ciências físicas. A noção de causalidade que está no centro dessas disciplinas envolve, principalmente, **relações determinísticas** – isto é, relações em que, se uma causa ocorre, então o efeito ocorrerá *com certeza*. Em contrapartida, porém, o mundo das interações humanas consiste em **relações probabilísticas** – nesse sentido, um aumento em *X* está associado ao aumento (ou à diminuição) na probabilidade de *Y* ocorrer, mas essas probabilidades não são certezas. Enquanto leis da física, como as leis de Newton, são determinísticas – pense na lei da gravidade –, as ciências sociais (incluindo a ciência política) as semelham-se mais ao nexo de causalidade probabilísticas da teoria da seleção natural de Darwin, na qual mutações aleatórias tornam um organismo mais ou menos capaz de sobreviver e se reproduzir<sup>4</sup>.

O que significa dizer, na ciência política, que nossas concepções de causalidade devem ser probabilísticas em sua natureza? Quando teorizamos, por exemplo, que o nível de riqueza de um indivíduo causa sua opinião sobre qual é política de tributação ótima, não queremos dizer que *toda* pessoa rica quererá impostos baixos e que *toda* pessoa pobre preferirá impostos altos. Considere o que aconteceria se encontrássemos uma única pessoa rica favorável a impostos altos ou uma única pessoa pobre favorável a impostos baixos. (Talvez você seja ou conheça tal pessoa.) Um caso sozinho não diminui a confiança em nossa teoria. Em vez de dizermos deterministicamente que “pessoas ricas preferirão impostos baixos e pessoas pobres preferirão impostos altos”, dizemos que “pessoas ricas têm maior probabilidade de preferir impostos baixos e pessoas pobres têm maior probabilidade de preferir impostos altos”. Nesse sentido, a relação é probabilística, não determinística.

Vejamos outro exemplo: estudiosos dos conflitos internacionais têm notado que existe uma relação estatística entre o tipo de regime de um país e a chance desse país entrar em guerra. Mais precisamente, em uma ampla série de estudos conhecida como literatura da “paz democrática”, muitos pesquisadores têm notado que guerras têm menor probabilidade de ocorrer entre dois países que são democracias do que entre pares de países em que ao menos um é uma não democracia. Para ser perfeitamente claro: a literatura não sugere que democracias nunca entram em guerra, mas que democracias não combatem outras democracias. Uma variedade de mecanismos tem sido sugerida para explicar

<sup>3</sup> Você pode encontrar uma excelente amostra do vigor desses debates em um livro de 2003 escrito por David Edmonds e John Eidinow, chamado *Wittgenstein's Poker: The Story of a Ten Minute Argument Between Two Great Philosophers*.

<sup>4</sup> Todavia, durante a revisão de três tentativas proeminentes de elaboração sobre a natureza probabilística da causalidade na filosofia da ciência, o filósofo Wesley Salmon (1993, p. 137) notou que “Na vasta literatura filosófica sobre causalidade [noções de causalidade probabilística] são largamente ignoradas”. Empréstamos de Brady (2004) essa útil comparação entre a natureza probabilística da ciência social e a teoria darwiniana da seleção natural.

essa correlação, mas o ponto aqui é que, se duas democracias começarem uma guerra entre si no próximo ano, seria um erro descartar a teoria. Uma teoria determinística diria que “democracias não entram em guerra entre si”, mas uma teoria probabilística mais sensível diz que “é altamente improvável que democracias entrem em guerra entre si”.

Na ciência política existirão sempre exceções, porque seres humanos não são robôs cujo comportamento sempre segue orientações preestabelecidas. Em outras ciências nas quais o objeto de estudo não possui livre-arbítrio, faz mais sentido falar em leis que descrevem comportamentos. Considere, por exemplo, o estudo das órbitas planetárias, em que cientistas podem prever o movimento dos corpos celestes centenas de anos antes de este ocorrer. O mundo político, em contrapartida, é extremamente difícil de prever. Como resultado, na maioria do tempo, estamos felizes por sermos capazes de fazer afirmações sobre relações causais probabilísticas.

O que tudo isso quer dizer, em resumo, é que a noção do que significa dizer que algo “causa” algo é uma questão que está longe de ser resolvida. Em face disso, os cientistas sociais devem abandonar a busca por conexões causais? De modo algum. O que isso significa é que devemos proceder com cuidado e com a mente sempre aberta, em vez de assumirmos um modo excessivamente rígido de pensar.

### 3.2 QUATRO OBSTÁCULOS NA ROTA PARA O ESTABELECIMENTO DE RELAÇÕES CAUSAIS

Se desejarmos investigar se alguma variável independente, que chamaremos de  $X$ , “causa” alguma variável dependente, que chamaremos de  $Y$ , quais procedimentos deveremos seguir antes que possamos expressar nosso grau de confiança se a relação causal existe ou não existe? Encontrar algum tipo de covariação (correlação) entre  $X$  e  $Y$  não é suficiente para uma conclusão desse tipo.

Estimulamos você a ter em mente que estabelecer relações causais entre variáveis não é de modo algum como caçar evidências com DNA, como ocorre nas séries criminais de televisão. A realidade social não apresenta respostas tão simples e diretas. Seguindo a discussão anterior sobre a natureza da causalidade, considere o que apresentaremos como diretrizes para o que constitui as “melhores práticas” da ciência política. Para qualquer teoria sobre uma relação causal entre  $X$  e  $Y$ , devemos cuidadosamente considerar as respostas para as quatro questões abaixo:

1. Existe algum mecanismo causal crível que conecta  $X$  a  $Y$ ?
2. Podemos eliminar a possibilidade de que  $Y$  pode causar  $X$ ?
3. Existe covariação entre  $X$  e  $Y$ ?
4. Controlamos por todas as **variáveis colineares**  $Z$  que podem tornar a associação entre  $X$  e  $Y$  **espúria**<sup>5</sup>?

<sup>5</sup> Uma “variável colinear” é simplesmente uma variável que é correlacionada com a variável independente e dependente e que de alguma forma altera a relação entre essas duas variáveis. “Espúria” significa “que não é o que aparenta ser” ou “falsa”.

Primeiro, devemos considerar se é crível afirmar que *X pode causar Y*. Com efeito, esse obstáculo representa um esforço de responder às questões “como” e “por que” em relação à teorias causais. Para fazer isso, precisamos realizar um exercício mental no qual avaliamos os mecanismos de como *X* pode causar *Y*. Qual é o processo ou o mecanismo que, logicamente falando, sugere que *X* pode ser a causa de *Y*? Em outras palavras, o que especificamente faz com que ter mais (ou menos) de *X* aumente a probabilidade de ter mais (ou menos) de *Y*? Quanto mais estranhos forem esses mecanismos, menos confiantes estaremos em que nossas teorias conseguem superar esse primeiro obstáculo. Falhar em superar de maneira clara esse primeiro obstáculo é um problema sério; o resultado é que ou nossa teoria precisa ser abandonada totalmente, ou precisamos revisá-la após pensarmos cuidadosamente sobre os mecanismos subjacentes pelos quais ela funciona. É fundamental que procedamos para a segunda pergunta somente quando tivermos um “sim” como resposta à primeira questão.

Segundo, devemos perguntar se podemos eliminar a possibilidade de que *Y* pode causar *X*. Como você aprenderá da discussão sobre as várias estratégias para avaliar conexões causais no capítulo 4, isso coloca problemas mais espinhosos para algumas áreas da pesquisa em ciências sociais do que para outras. Ocasionalmente, esse obstáculo pode ser superado por meio da lógica. Por exemplo, quando consideramos se o gênero de uma pessoa (*X*) faz com que ele ou ela tenha atitudes específicas sobre a política de aborto (*Y*), fica óbvio que o cenário de causalidade reversa pode ser desconsiderado: as atitudes de uma pessoa sobre o aborto não “causam” que ela seja homem ou mulher. Se nossa teoria não consegue claramente superar esse obstáculo específico, a corrida não está perdida. Isto é, não precisamos abandonar nossa teoria. Sob essas circunstâncias, devemos proceder para a próxima pergunta, sem deixar de ter em mente a possibilidade de que a nossa relação causal possa ser revertida.

Durante nossas considerações sobre os dois primeiros obstáculos para uma teoria causal, estivemos preocupados apenas com duas variáveis, *X* e *Y*. O terceiro obstáculo pode envolver uma terceira variável *Z* e o quarto obstáculo sempre a envolve. Frequentemente temos várias variáveis *Z*.

No terceiro obstáculo para uma teoria causal, devemos considerar se *X* e *Y* covariam (ou, em outros termos, se elas são correlacionadas ou associadas). Falando de maneira geral, para *X* causar *Y*, deve existir alguma forma mensurável de associação entre *X* e *Y*, por exemplo, “a maior presença de *X* está associada a uma maior presença de *Y*” ou “uma maior presença de *X* está associada a uma menor presença de *Y*”. Demonstrar uma conexão bivariada simples é uma questão direta (cobriremos esse tópico nos capítulos 7 e 8). Claro, você pode estar familiarizado com o ditado “Correlação não prova causalidade”. Concordamos totalmente com ele. Porém, é importante notar que correlação é normalmente um componente essencial para a causalidade. Mas é importante ser cuidadoso. É possível que exista uma relação causal entre *X* e *Y* mesmo se não existir uma associação bivariada entre *X* e *Y*. Assim, mesmo que falhemos em superar esse obstáculo, não devemos abandonar totalmente nossa afirmação causal. Pelo contrário, devemos considerar a possibilidade de que exista alguma variável colinear *Z* pela qual precisamos “controlar” antes de observar a relação entre *X* e *Y*.

Devemos prosseguir para o nosso quarto e final obstáculo, encontrando ou não uma relação entre  $X$  e  $Y$ .

Quarto, para estabelecer uma conexão causal entre  $X$  e  $Y$ , devemos enfrentar a realidade de que, como notamos no início deste capítulo, vivemos em um mundo no qual muitas das variáveis dependentes são causadas por mais de uma variável. Quais são os problemas que isso apresenta para as ciências sociais? Isso significa que, quando tentamos estabelecer se um  $X$  específico causa um  $Y$  específico, precisamos “controlar” efeitos das outras causas de  $Y$  (podemos chamar esses outros efeitos de  $Z$ ). Se falharmos em controlar pelos efeitos de  $Z$ , é bastante provável que entendamos equivocadamente a relação entre  $X$  e  $Y$  e que façamos inferências erradas sobre  $X$  causar  $Y$ . Esse é o erro mais sério que um cientista social pode fazer. Se encontrarmos que  $X$  e  $Y$  são correlacionados, mas que, quando controlamos pelos efeitos de  $Z$  em  $X$  e  $Y$ , a associação entre  $X$  e  $Y$  desaparece, então a relação entre  $X$  e  $Y$  é considerada espúria.

### 3.2.1 CONSIDERANDO TUDO JUNTO – ADICIONANDO AS RESPOSTAS PARA NOSSAS QUATRO PERGUNTAS

Como acabamos de ver, o processo de avaliação de uma afirmação teórica de que  $X$  causa  $Y$  é complicado. Cada uma das quatro questões que introduzimos nesta seção pode ser difícil de ser respondida com clareza. Mas o desafio de avaliar uma afirmação de que  $X$  causa  $Y$  envolve a soma das respostas de todas as quatro perguntas. Só assim podemos determinar nossa confiança geral sobre  $X$  causar  $Y$ . Para entender isso, pense sobre a analogia que temos usado de chamar essas perguntas de “obstáculos”. Em uma corrida de obstáculos, os corredores devem fazer o seu melhor para superar cada um deles enquanto correm em direção à linha de chegada. Ocasionalmente, mesmo os mais experientes irão derrubar um ou outro. Embora isso possa diminuir sua velocidade e sua chance de vitória, nem tudo está perdido. Se pensarmos em avaliar se uma teoria consegue superar os quatro obstáculos colocados pelas questões precedentes, não há dúvida de que nossa confiança aumentará quando formos capazes de responder às quatro perguntas de modo direto e sem reservas (“sim”, “sim”, “sim” e “sim”). Como descrevemos na introdução desta seção, falhar em superar o primeiro dos obstáculos deve nos fazer parar e repensar nossa teoria. Também devemos fazer isso se acharmos que nossa relação é espúria. Para o segundo e terceiro obstáculos, todavia, falhar em superá-los completamente não significa que devemos descartar nossa afirmação causal. A Figura 3.1 apresenta um sumário desse processo. Na próxima seção, apresentaremos todo o processo descrito na Figura 3.1 utilizando uma série de exemplos.

Conforme percorrermos este processo de responder às quatro perguntas, manteremos entre colchetes um **scoreboard dos obstáculos causais** como um atalho para sumarizar as respostas. Por agora, limitaremos nossas respostas a “s” para “sim”, “n” para “não” e “?” para “talvez”. Se uma teoria tiver superado os quatro obstáculos, seu

scoreboard de resultados deverá conter [s s s] e a afirmação causal subjacente terá um forte apoio. Como descrevemos acima, esses obstáculos não possuem os mesmos impactos sobre nossa afirmação de causalidade. Então, por exemplo, uma afirmação causal que tiver o placar [n s s] poderia ser instantaneamente descartada. Mas uma afirmação que tiver um com [s n s] tem um nível razoável de evidências em seu favor.

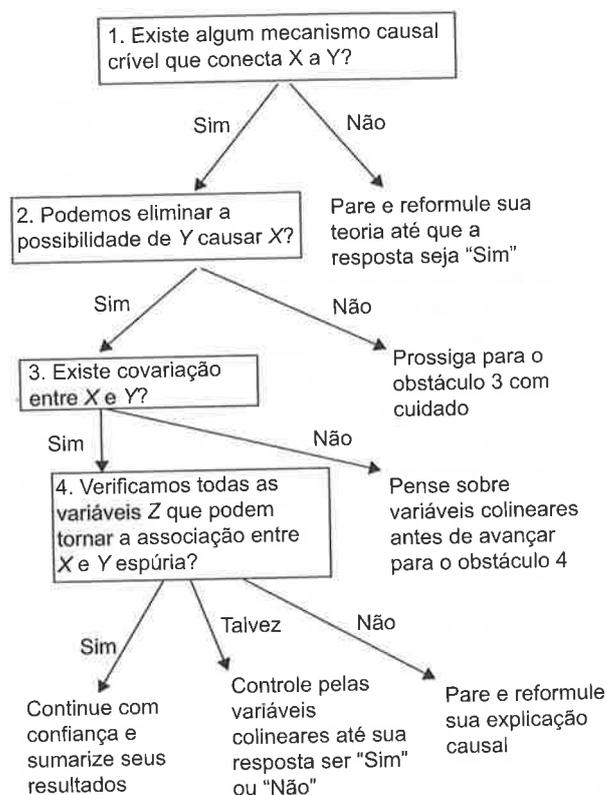


Figura 3.1 – O percurso para avaliar uma relação causal.

### 3.2.2 IDENTIFICAR AFIRMAÇÕES CAUSAIS É UMA ATIVIDADE MENTAL ESSENCIAL

Queremos enfatizar que a lógica apresentada anteriormente não se aplica meramente a exemplos de pesquisa em ciência política. Todas as vezes que você vir uma história nos jornais ou ouvir um discurso de um candidato a um cargo público, ou ler um artigo para uma aula de ciência política, quase sempre haverá uma afirmação causal embutida na história, no discurso ou no artigo. Algumas vezes essas afirmações causais estão explícitas, o que evita que você não as perceba. Porém, muito frequente-

mente, elas são difíceis de serem achadas, e, na maior parte do tempo, não porque o orador ou o escritor está tentando confundir-lo. O que queremos enfatizar é que localizar e identificar afirmações causais é uma habilidade mental. Isso não é natural para a maioria das pessoas, mas pode ser praticado.

No nosso dia a dia, somos frequentemente apresentados a afirmações causais de pessoas que nos tentam persuadir para que adotemos o ponto de vista delas. Advogar e tentar persuadir são obviamente componentes saudáveis de uma democracia saudável. O debate público, contudo, será ainda mais saudável quando cidadãos escrutinarem ativamente as afirmações com as quais são confrontados. Pegue, por exemplo, os debates na mídia sobre os méritos dos *private school choice programs*<sup>6</sup>, que tem sido implementado em várias escolas. Entre os argumentos a favor de tais programas estão aqueles que afirmam que eles melhoraram o desempenho dos estudantes nos testes padronizados. Relatórios da mídia sobre os sucessos e os fracassos desse tipo de programa são bastante comuns. Por exemplo, observe como um artigo no *Washington Post* expõe o argumento de um artigo acadêmico sobre um programa:

Estudantes afro-americanos no Distrito [de Columbia] e outras duas cidades melhoraram seu desempenho em comparação com seus colegas da escola pública após terem sido transferidos para escolas privadas com o auxílio de *vouchers*<sup>7</sup>, de acordo com um novo estudo. O estudo mostra que os alunos transferidos para escolas privadas obtiveram resultados mais altos em seis pontos percentuais do que os que foram mantidos em escolas públicas na Cidade de Nova York, Dayton, Ohio e no Distrito. O maior efeito foi observado no Distrito, onde estudantes participantes do programa foram nove pontos percentuais melhores do que seus colegas de escolas públicas<sup>8</sup>.

Observe a afirmação causal feita aqui: participar (ou não) de um programa de escolha de escola (*X*) faz com que o resultado de uma criança em um teste (*Y*) varie. Frequentemente, o leitor é apresentado a algum gráfico de barras que apoia o argumento. Ele é encorajado a pensar, às vezes sutilmente, que o diferente tamanho das barras representam diferentes médias para o resultado do teste de uma criança que participa do programa e outra da escola pública, significando que o programa é a causa dos melhores resultados das crianças do programa. Quando somos confrontados com esse tipo de informação, podemos ser tentados a pegar essa pequena evidência e concluir que existe uma relação causal. A principal lição que se deve aprender aqui é que essa é uma conclusão prematura.

Vamos ser claros: programas de escolha de escola podem, de fato, fazer com que estudantes se saiam melhor em testes padronizados. Nosso objetivo não é avançar nesse debate, mas mostrar como são necessárias habilidades cognitivas para avaliar as afirmações

<sup>6</sup> Nos Estados Unidos, programas que permitem que os pais tenham o direito de escolher como utilizar recursos fornecidos pelo Estado para financiar a educação de ensino fundamental e médio de seus filhos.

<sup>7</sup> Nos Estados Unidos, os participantes de tal programa recebem um comprovante (*voucher*) que pode ser utilizado para o pagamento da educação em escolas privadas com recursos públicos. [N.T.]

<sup>8</sup> Mathews, Jay. Scores Improve for D.C. Pupils With Vouchers. *Washington Post*, A1, 28 ago. 2000.

causais feitas pelos defensores e pelos críticos desses programas. Apresentar evidências de que estudantes que participam desses programas possuem melhores resultados do que estudantes que pertencem à escola pública é somente *uma peça* do quebra-cabeça causal – nominalmente, isso faz com que o terceiro obstáculo seja superado: existe covariação entre  $X$  e  $Y$ . Até o momento, em nossa avaliação, superamos apenas uma das barreiras. Nosso *scoreboard* de resultado é  $[? \ ? \ ?]$ . Assim, antes de concluirmos que programas de escolha de escolas causam (ou não) o desempenho do estudante, precisamos submeter essa afirmação aos quatro obstáculos, não apenas ao terceiro deles.

Então vamos submeter a afirmação aos quatro obstáculos. Primeiro, existe um mecanismo que podemos usar para explicar como e por que frequentar um tipo de escola específico – pública ou privada paga por meio do programa – pode afetar o resultado de um estudante em um teste? Certamente. Muitas escolas privadas que participam desse tipo de programa têm turmas menores (entre outros benefícios), e turmas menores podem se traduzir em maior aprendizado e notas mais altas. *Portanto, a resposta para a primeira pergunta é “sim”*  $[s \ ? \ s \ ?]$ . Segundo, é possível que a relação causal ocorra no sentido contrário? Isto é, podemos eliminar a possibilidade de que os resultados dos testes façam com que uma pessoa participe ou não participe de um programa de escolha de escolas? Como os testes acontecem meses ou mesmo anos após a pessoa escolher a escola, não é possível que os resultados determinem a participação no programa. *Assim, a resposta para a segunda questão também é “sim”*  $[s \ s \ s \ ?]$ . Terceiro, existe uma correlação entre participar de um programa e os resultados em testes? O artigo citado acima observou que, nas três cidades consideradas, existe – estudantes participantes do programa possuem notas melhores em testes padronizados do que seus colegas de escolas públicas. *A resposta à terceira pergunta é, portanto, “sim”*  $[s \ s \ s \ ?]$ . Finalmente, controlamos por todas as variáveis colineares que podem fazer com que a associação entre a participação no programa e o resultado dos testes seja espúria? Lembre-se, uma potencial variável colinear é simplesmente uma variável que está relacionada à variável independente e que também é uma causa da variável dependente. Podemos pensar em algo que pode estar relacionado ao tipo de escola que uma criança frequenta e que também pode ser uma provável causa do desempenho da criança em um teste? Claro que podemos. A variável “envolvimento dos pais” é uma candidata natural a ser uma variável  $Z$  neste caso. Algumas crianças têm pais altamente envolvidos – pais que leem para elas, que as ajudam na lição de casa e que possuem um papel ativo na educação delas –, enquanto outras crianças possuem pais que não são tão envolvidos. Pais altamente envolvidos têm maior probabilidade de saber sobre a existência de programas de escolha de escola em suas cidades e de candidatar seus filhos a esses programas. (Então é praticamente certeza que  $Z$  está relacionado a  $X$ .) E pais altamente envolvidos também têm maior probabilidade de criar altas expectativas nos filhos e de inculcar em seus filhos a percepção de que o sucesso na escola é importante, o que, tomado em conjunto, pode se traduzir em um melhor desempenho da criança nos testes padronizados. (Então  $Z$  deve ser a causa de  $Y$ .) A questão principal, então, se torna: o estudo em questão *controlou* por esses outros efeitos? Estamos um pouco adiantados no conteúdo neste momento, porque ainda não falamos sobre as estratégias que os pesquisadores utilizam para controlar pelo efeito de

variáveis colineares potenciais. (Esse assunto é tratado no capítulo 4.) Mas esperamos que você possa observar por que controlar pelo efeito do envolvimento dos pais é tão importante nesta situação em particular (e em geral): se nossa comparação entre uma criança de uma escola do programa e uma criança de uma escola pública equivale, basicamente, a uma comparação entre uma criança com pais altamente envolvidos e uma com pais parcamente envolvidos, então se torna muito problemático concluir que a diferença entre o resultado dos alunos é *causado* pelo programa. Sem controlar pelo envolvimento dos pais ( $Z$ ), em outras palavras, a relação entre o tipo de escola ( $X$ ) e o resultado dos testes ( $Y$ ) pode ser espúria. Então, até que vejamos evidências de que esse importante  $Z$  tenha sido utilizado como controle, a afirmação causal não terá passado pelos quatro obstáculos, nosso *scoreboard* de resultados é  $[s s s n]$  e devemos suspeitar altamente dos achados do estudo. De modo mais informal, sem tal controle, a comparação entre esses resultados dos testes dos dois grupos é injusto, porque os grupos podem ser inicialmente diferentes. No artigo do *Washington Post* que mencionamos, o envolvimento dos pais é controlado, uma vez que os estudantes foram escolhidos para o programa de maneira aleatória. Apresentaremos apenas no capítulo 4 por que isso faz diferença, mas, acredite, faz.

O mesmo processo pode ser aplicado a uma ampla variedade de afirmações causais e perguntas com as quais nos deparamos no nosso dia a dia. Beber vinho tinto causa uma redução nas doenças cardíacas? Psicoterapia ajuda as pessoas com problemas emocionais e de relacionamento? Aumentar o gasto do governo estimula ou retarda o crescimento econômico? Em cada um desses casos e em muitos outros exemplos, podemos ser tentados a observar correlação entre duas variáveis e concluir que a relação é causal. É importante resistirmos a essa tentação e submetermos cada uma dessas afirmações aos critérios mais rigorosos que estamos sugerindo. Se pensarmos sobre nossa evidência nos termos do nosso *scoreboard* de resultados, o que temos é  $[? ? s ?]$ . Esse é um começo razoável para a avaliação de uma afirmação causal, mas é um lugar muito problemático para se encerrar e extrair conclusões definitivas. Pensando nos termos da superação dos quatro obstáculos, quando alguém nos apresenta uma afirmação causal que falha em superar cada um deles, nós naturalmente faremos perguntas adicionais e, quando fizermos isso, nos tornaremos consumidores de informações mais inteligentes na nossa vida cotidiana.

Uma parte importante de uma abordagem científica para o estudo da política é que sejamos céticos quanto às afirmações de outros estudiosos sobre relações causais. Antes de podermos avaliar a teoria causal, precisamos considerar quão bem as evidências disponíveis respondem a cada uma das quatro questões sobre  $X$ ,  $Y$  e  $Z$ . Uma vez que tenhamos respondido cada uma das quatro questões, podemos pensar sobre o nível de confiança geral que temos na afirmação de que  $X$  causa  $Y$ .

### 3.2.3 QUAIS SÃO AS CONSEQUÊNCIAS DE FALHAR NO CONTROLE DE OUTRAS POSSÍVEIS CAUSAS?

Quando se trata de afirmações causais, como acabamos de observar, é no quarto obstáculo causal que frequentemente tropeçamos, e isso não acontece apenas na re-

tórica política e nas histórias apresentadas na mídia. Isso também é verdade para o escrutínio das pesquisas científicas. De fato, uma parte substancial dos desacordos entre acadêmicos tem origem no quarto obstáculo causal. A objeção mais frequente quando um acadêmico está avaliando o trabalho de outro acadêmico talvez seja que um pesquisador “falhou em controlar por” alguma causa potencialmente importante da variável dependente.

O que acontece quando falhamos em controlar por alguma outra causa plausível da nossa variável dependente de interesse? Essa resposta é bastante simples: significa que falhamos em superar o quarto obstáculo causal. *Enquanto existir um argumento razoável sobre como uma variável Z não controlada pode estar relacionada com X e Y, não podemos concluir com total confiança que X de fato causa Y.* Como o principal objetivo da ciência é estabelecer se existem relações causais entre variáveis, falhar em controlar por outras causas de Y é um problema potencialmente sério.

Um dos temas deste livro é que a análise estatística não deve ser desconectada das questões do desenho de pesquisa – como controlar por quantas causas da variável dependente forem possíveis. Quando discutirmos o modelo de regressão multivariado (nos capítulos 9, 10 e 11), que é a técnica estatística mais comum nas pesquisas de ciência política, a argumentação será centrada em aprender a controlar por outras causas da variável dependente. Veremos que uma falha do desenho de pesquisa, como falhar em controlar por todas as causas relevantes da variável dependente, possuem implicações estatísticas, e que as implicações são sempre ruins. Mantenha em mente que falhas do desenho de pesquisa produzem problemas na análise estatística. Por agora, o importante é que bons desenhos de pesquisa dão mais credibilidade à análise estatística, enquanto desenhos de pesquisa ruins fazem que seja difícil para qualquer análise estatística ser conclusiva sobre conexões causais.

### 3.3 POR QUE O ESTUDO DA CAUSALIDADE É TÃO IMPORTANTE? TRÊS EXEMPLOS DA CIÊNCIA POLÍTICA

Nossa ênfase nas conexões causais deve estar clara. Voltamos agora a uma série de controvérsias correntes dentro da disciplina de ciência política para mostrar como o debate sobre causalidade está no centro das controvérsias que fazem você (e a maioria de nós) ter interesse em política.

#### 3.3.1 SATISFAÇÃO COM A VIDA E A ESTABILIDADE DEMOCRÁTICA

Uma das controvérsias mais duradouras na ciência política é a relação entre *satisfação pública com a vida e a estabilidade das instituições democráticas*. Satisfação com a vida pode, obviamente, significar diferentes coisas, mas nesta discussão consideramos que ela varia ao longo de um contínuo, da alta insatisfação com a vida do dia a dia à alta satisfação. Qual é (se é que existe alguma) a conexão causal entre esses dois conceitos?

O cientista político Ronald Inglehart (1988) argumenta que a satisfação com a vida (X) causa a estabilidade de um sistema democrático (Y). Se pensarmos na primeira das nossas quatro questões sobre o estabelecimento de relações causais, podemos observar que existe um mecanismo causal crível que conecta X a Y – se as pessoas em um país democrático estão mais felizes com suas vidas, elas têm menor probabilidade de querer derrubar o governo. A resposta para a primeira pergunta então é “sim” [s ? ? ?]. Passando para a segunda pergunta: podemos eliminar a possibilidade da estabilidade democrática (Y) ser a causa da satisfação com a vida (X)? Não podemos. É bastante fácil apresentar um mecanismo causal que proponha que cidadãos de democracias estáveis têm maior probabilidade de estarem satisfeitos com suas vidas do que cidadãos de países com uma história de instabilidade e com governos não democráticos. Portanto, a resposta à segunda pergunta é “não” [s n ? ?]. Agora, nos debruçamos sobre a terceira questão. Utilizando uma quantidade impressionante de dados para uma ampla variedade de países desenvolvidos, Inglehart e seus colegas têm mostrado que existe, de fato, uma associação entre o nível médio da satisfação pública e a quantidade de anos ininterruptos de governo democrático. Isto é, países com médias altas de satisfação com a vida têm desfrutado de períodos ininterruptos mais longos de estabilidade democrática. Por outro lado, países com níveis menores de satisfação com a vida têm tido períodos menores de estabilidade democrática e um maior número de agitações revolucionárias. A resposta à nossa terceira pergunta é, portanto, “sim” [s n s ?]. No que diz respeito à quarta questão, é fácil imaginar uma miríade de outros fatores (Zs) que levam à estabilidade democrática, e se Inglehart tem feito um trabalho adequado em controlar por esses outros fatores é assunto de considerável debate acadêmico. A resposta para essa quarta pergunta é “talvez” [s n s ?]. A teoria de Inglehart responde satisfatoriamente às perguntas 1 e 3, mas sua resposta às perguntas 2 e 4 é que tem dado aos céticos razões substanciais para duvidar de sua afirmação causal.

### 3.3.2 RAÇA E PARTICIPAÇÃO POLÍTICA NOS ESTADOS UNIDOS

Participação política – o quanto cada cidadão se engaja em atividades políticas voluntárias, como votar, trabalhar em campanhas ou contribuir financeiramente – representa uma das facetas mais frequentes dos estudos de comportamento de massas, especialmente nos EUA. Isso ocorre por uma boa razão: a participação em sociedades democráticas é vista por alguns como uma medida da saúde de uma democracia. Após décadas de estudos sobre a variação das taxas de participação dos americanos, diversas características demográficas consistentemente se destacaram por estarem correlacionadas com a participação, incluindo a autoclassificação racial. *Surveys* têm constantemente mostrado que a participação de caucasianos na política é consideravelmente mais frequente do que a de latinos ou afro-americanos. Um *survey* de grande alcance, por exemplo, mostra que, durante um ciclo eleitoral típico, caucasianos participam de 2,22 “atos” – tais como votar, trabalhar em campanhas, fazer contribuição a campanhas, participar de protestos – enquanto as taxas de afro-americanos e de latinos são de 1,90 e 1,41 atividade (ver Verba *et al.*, 1993, Figura 1).

A relação entre a raça do indivíduo ( $X$ ) e o quanto ele participa da política ( $Y$ ) é causal? Antes de aceitarmos as evidências apresentadas como conclusivas da existência de uma relação *causal*, precisamos submeter essa afirmação aos quatro obstáculos causais. Existe um mecanismo razoável que responda “como” e “por que” participação política e raça estão conectadas? Pode existir uma razão para pensarmos isso. Afinal, por um longo período na história americana, algumas barreiras formais e muitas informais existiram proibindo ou desencorajando a participação de não caucasianos. Supor que podem ainda existir efeitos residuais dessas barreiras, mesmo décadas após sua erradicação, é inteiramente razoável. *A resposta à primeira pergunta é “sim” [s ? ? ?].* Podemos eliminar a possibilidade de que a variação das taxas de participação cause a classificação racial do indivíduo? Obviamente que sim. *A resposta à nossa segunda pergunta é “sim” [s s ? ?].* Existe correlação entre a raça do indivíduo e seu nível de participação nos EUA? Os dados citados acima sobre a participação em atos de caucasianos, afro-americanos e latinos claramente mostram que existe uma relação; caucasianos participam mais. *A resposta à nossa terceira pergunta é “sim” [s s s ?].* Finalmente, controlamos por todas as possíveis variáveis colineares  $Z$  que são relacionadas com raça ( $X$ ) e participação ( $Y$ ) que podem fazer a relação ser espúria? Verba e seus colegas sugerem que pode existir apenas uma variável colinear: o *status* socioeconômico. Menos hoje do que no passado, o *status* socioeconômico ( $Z$ ) pode ser apontado como uma causa da participação política ( $Y$ ); pessoas ricas doam mais, se voluntariam mais etc. que sua contraparte menos rica. Uma vez que se controla pelo *status* socioeconômico, a relação entre raça e participação política desaparece por inteiro (ver Verba *et al.*, 1993, Tabela 8). Em resumo, a correlação que observamos entre raça e participação política é espúria (ou ilusória); não ocorre em função da raça, mas em função de disparidades de riqueza entre caucasianos e outras raças. Uma vez que controlamos por essas diferenças socioeconômicas, a conexão entre raça e participação some. *Portanto, a resposta para nossa quarta pergunta é “não” [s s s n].* Nesse caso, o esforço feito para responder à quarta pergunta muda nossa resposta para a terceira. Mudamos de [s s s ?] para [s s n n]. Esse é um dos modos importantes pelos quais nossas conclusões sobre relações podem mudar quando passamos de uma análise bivariada, na qual medimos a relação entre uma variável independente ( $X$ ) e uma variável dependente ( $Y$ ), para uma análise multivariada, na qual medimos a relação entre  $X$  e  $Y$  controlando por uma segunda variável independente ( $Z$ ). Também é possível que muitas outras coisas aconteçam quando controlamos por  $Z$ . Por exemplo, também é possível que nossas respostas mudem de [s s n n] para [s s s s].

### 3.3.3 AVALIANDO SE O PROGRAMA HEAD START É EFETIVO

Na década de 1960, como parte da guerra à pobreza<sup>9</sup>, o presidente Lyndon Johnson iniciou o programa Head Start para propiciar às crianças não economicamente

<sup>9</sup> Guerra à pobreza (*war on poverty*) foi uma série de iniciativas do presidente americano Lyndon Johnson (1963-1969) para combater a pobreza no país. [N.T.]

privilegiadas uma experiência de pré-escola que – o programa esperava – aumentaria as chances de crianças pobres serem bem-sucedidas uma vez que chegassem ao pré-primário e às demais séries. O programa claramente possuía boas intenções, mas, claro, apenas isso não o tornaria efetivo. Colocando de maneira simples: o programa funciona? Nesse caso, “funcionar” significa que o Head Start aumentaria as chances dos participantes do programa de ter um melhor resultado educacional do que os não participantes.

É tentador fazer a comparação simples dos resultados em testes padronizados de crianças que participaram do Head Start com aquelas que não participaram. Se os participantes do Head Start possuírem melhores notas, então – *voilà* – caso encerrado: o programa funciona. Caso contrário, então não funciona. Mas, como antes, precisamos nos manter focados nos quatro obstáculos causais. Primeiro, existe algum mecanismo causal crível que possa responder “como” e “por que” a participação no Head Start ( $X$ ) e os resultados educacionais ( $Y$ ) estão conectados? Sim, a teoria por detrás do programa é que a exposição ao ambiente pré-escolar que antecipa o ambiente escolar real ajuda a preparar a criança para o que ela encontrará no pré-primário e nos estágios posteriores. O Head Start, nesse sentido, pode ajudar a reduzir problemas de disciplina e preparar os estudantes para ler e contar, entre outras habilidades. *A resposta para nossa primeira pergunta é “sim” [s ? ? ?].* Segundo, é possível que a relação causal seja inversa? Em outras palavras, podemos eliminar a possibilidade de que os resultados educacionais ( $Y$ ) possam causar a participação no Head Start ( $X$ )? Como os testes são realizados anos após a participação no programa, então, sim. *A resposta para nossa segunda pergunta é “sim” [s s ? ?].* Existe alguma associação entre a participação no programa e os resultados de aprendizados? Estudo após estudo têm mostrado que participantes do Head Start se saem melhor quando são testados, e que com menor frequência repetem uma série do que os que não tiveram uma experiência de pré-escola. Por exemplo, um estudo amplamente citado mostra que crianças do Head Start se saem melhor em testes de vocabulário adequados para crianças pequenas do que alunos que não frequentaram a pré-escola (Currie e Thomas, 1995). *A resposta à nossa terceira pergunta, portanto, é “sim” [s s s ?].*

Como no caso dos exemplos dos programas escolares anteriormente mencionados, uma possível variável colinear é o envolvimento dos pais ( $Z$ ). Pais altamente envolvidos ( $Z$ ) têm maior probabilidade de buscar, ter conhecimento sobre, e inscrever seus filhos ( $X$ ) em programas como o Head Start, que podem beneficiá-los. Pais menos envolvidos com a vida dos filhos têm menor probabilidade de beneficiá-los com as potenciais oportunidades criadas pelo Head Start. E, como antes, é provável que o alto envolvimento dos pais ( $Z$ ) tenha efeitos positivos nos resultados educacionais dos filhos. A pergunta-chave, então, se torna: o envolvimento dos pais ( $Z$ ) faz com que a relação entre o Head Start e os resultados educacionais futuros seja espúria? O estudo anteriormente mencionado de Currie e Thomas utiliza tanto controles estatísticos como controles no desenho de pesquisa para controlar pelo efeito dos pais e encontram que o efeito do Head Start sobrevive entre as crianças caucasianas, mas não entre as crianças afro-americanas (ver Tabela 4 da obra citada). Novamente, o termo “controles estatísticos” pode não ser tão claro quanto será posteriormente. Por ora,

é suficiente dizer que esses pesquisadores utilizaram todas as técnicas que estavam disponíveis para mostrar que o Head Start, de fato, produz um efeito positivo em algumas, mas não em todas as crianças. *A resposta à nossa quarta pergunta é um altamente qualificado “sim” [s s s s].*

### 3.4 ENCERRANDO

O aprendizado das habilidades cognitivas requeridas para avaliar afirmações causais requer prática. Elas são hábitos intelectuais que, como uma boa faca, ficam mais afiadas com o uso.

Como traduzir essas habilidades cognitivas em novos desenhos de pesquisa que buscam responder perguntas causais é o assunto do capítulo 4. Todos os “desenhos de pesquisa” que você aprenderá no próximo capítulo estão fortemente ligados aos problemas da avaliação de afirmações causais. À medida que prosseguimos, é essencial que você mantenha as lições deste capítulo em mente para tornar-se um melhor consumidor de informação, bem como para prosseguir no caminho de ser um produtor de pesquisas.

### CONCEITOS INTRODUZIDOS NESTE CAPÍTULO

- Bivariado – que envolve apenas duas variáveis.
- Espúrio – que não é o que aparenta ser, ou falso.
- Multivariado – que envolve mais que duas variáveis.
- Relação determinística – se alguma causa ocorre, então o efeito acontecerá com certeza.
- Relação probabilística – o aumento de  $X$  está associado ao aumento (ou decréscimo) na probabilidade de  $Y$  ocorrer, mas essas probabilidades não são certezas.
- *Scoreboard* dos obstáculos causais – abreviação para sumarizar evidências sobre a possível relação causal entre uma variável independente e uma variável dependente.
- Variável colinear – uma variável que é correlacionada com a variável independente e a variável dependente e que de algum modo pode alterar a relação entre essas duas variáveis.

### EXERCÍCIOS

1. Pense em uma aula de história na qual você aprendeu as “causas” de um evento histórico (por exemplo, a Grande Depressão, a Revolução Francesa ou a Primeira Guerra Mundial). Quão bem cada uma dessas afirmações causais se sai quan-

do você tenta responder a cada uma das quatro questões para o estabelecimento de relações causais?

2. Entre no *site* de um jornal de notícias. Na caixa de “pesquisa” do *site*, digite as palavras “pesquisa causal” (sem aspas). (Dica: talvez você também tenha que limitar o período temporal da busca, dependendo do *site* que você utilizar.) Nos resultados da pesquisa, encontre dois artigos que façam afirmações sobre relações causais. Faça uma breve sinopse da afirmação causal feita em cada um deles.
3. Imagine, para cada um dos seguintes exemplos, que algum pesquisador tenha achado o padrão de covariação entre  $X$  e  $Y$  que é descrito. Você consegue pensar em uma variável  $Z$  que pode fazer com que a relação entre  $X$  e  $Y$  seja espúria?
  - a) Quanto mais bombeiros ( $X$ ) atendem a um incêndio, maior é o dano que ocorre na propriedade ( $Y$ ).
  - b) Quando mais dinheiro é gasto por um congressista durante a campanha ( $X$ ), menor é seu percentual de votos ( $Y$ ).
  - c) O aumento no consumo de café ( $X$ ) reduz o risco de depressão entre as mulheres ( $Y$ ).
  - d) Quanto maior são os salários dos ministros presbiterianos ( $X$ ), maior é o preço do rum em Havana ( $Y$ ).
4. Escreva uma relação probabilística e uma determinística para descrever a relação em cada um dos seguintes pares de variáveis dependente e independente:
  - a) A educação de uma pessoa ( $X$ ) e o comparecimento eleitoral ( $Y$ ).
  - b) A saúde econômica de um país ( $X$ ) e as revoluções políticas ( $Y$ ).
  - c) A altura de um candidato ( $X$ ) e o resultado eleitoral ( $Y$ ).
5. Dê uma olhada no livro de códigos do banco de dados “BES 2005 Subset” e escreva suas respostas para os seguintes itens:
  - a) Desenvolva uma teoria causal sobre a relação entre uma variável independente ( $X$ ) e uma variável dependente ( $Y$ ) desse banco de dados. Existe um mecanismo causal crível que conecte  $X$  a ( $Y$ )? Explique sua resposta.
  - b) Pode  $Y$  causar  $Y$ ? Explique sua resposta.
  - c) Por quais outras variáveis ( $Z$ ) você gostaria de controlar nos testes de sua teoria?
6. Imagine afirmações causais para cada um dos *scoreboards* de resultados listados abaixo. Qual destas afirmações você avaliaria como a com maior suporte? Justifique sua resposta.
  - a)  $[s n s n]$
  - b)  $[s s s n]$
  - c)  $[? s s s]$

7. O pesquisador *A* e o pesquisador *B* estão tendo um debate científico. Sobre o que eles estão argumentando se a discussão for focada no:
  - a) obstáculo causal 1.
  - b) obstáculo causal 2.
  - c) obstáculo causal 3.
  - d) obstáculo causal 4.
8. Encontre um artigo em um periódico de ciência política que contenha um modelo sobre política. Cite a referência completa e a passagem do artigo, e depois:
  - a) Descreva brevemente a teoria causal que conecta as variáveis independente e dependente.
  - b) Crie um *scoreboard* de resultados dos obstáculos causais para essa teoria e escreva uma explicação para cada uma das suas respostas às quatro perguntas causais.