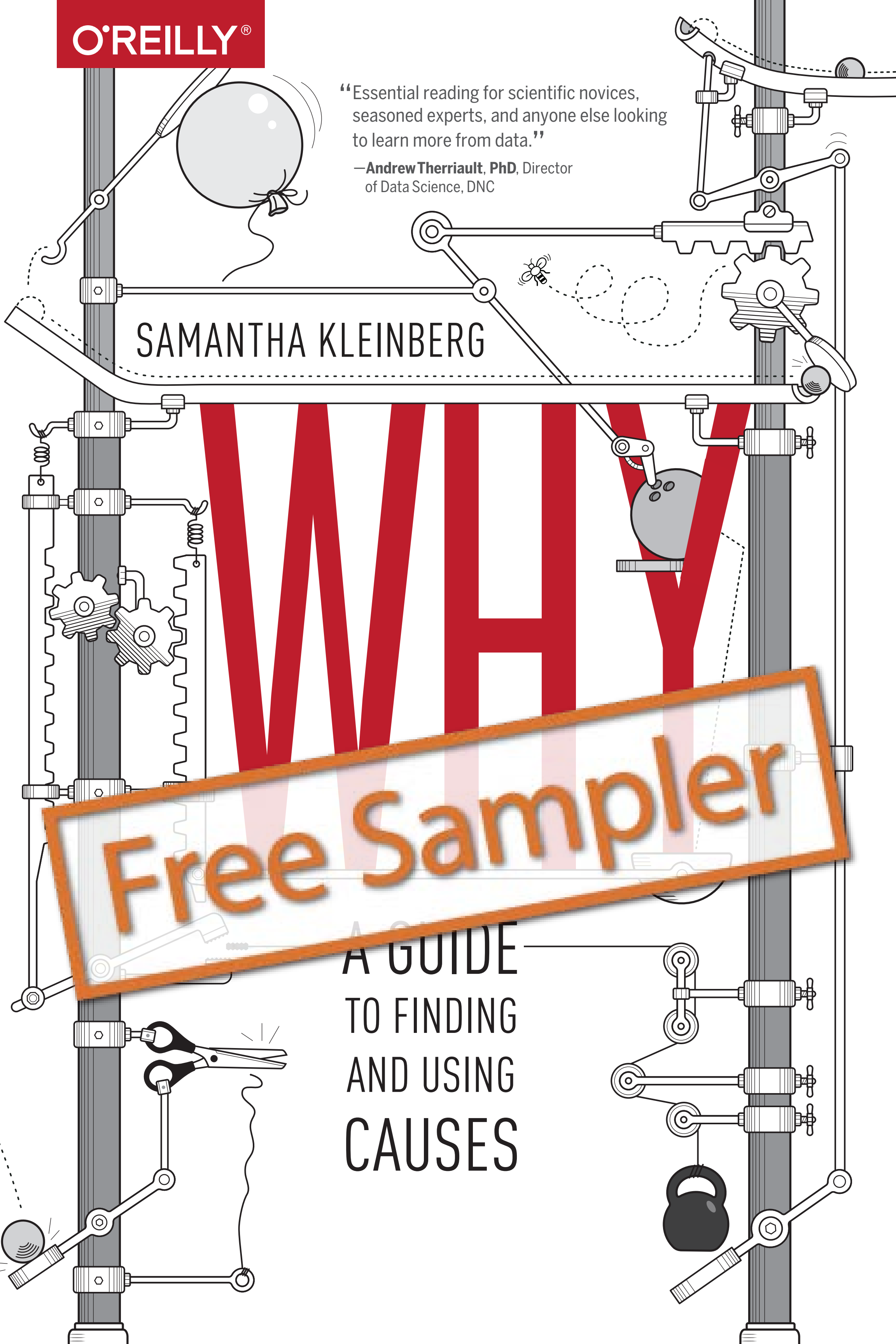


O'REILLY®

"Essential reading for scientific novices, seasoned experts, and anyone else looking to learn more from data."

—**Andrew Therriault, PhD**, Director of Data Science, DNC

SAMANTHA KLEINBERG

WHY

A GUIDE TO FINDING AND USING CAUSES

Free Sampler

O'REILLY®

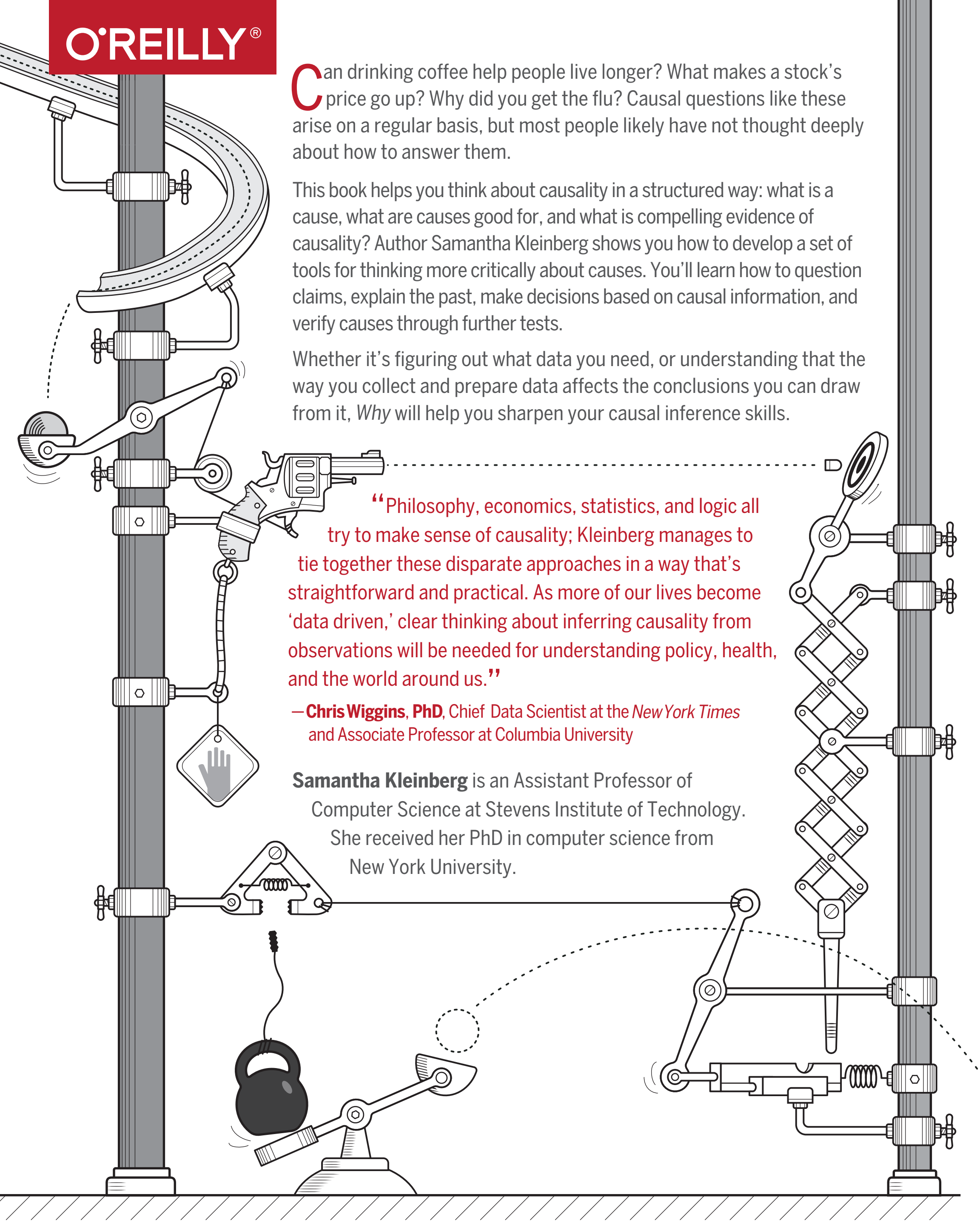
Can drinking coffee help people live longer? What makes a stock's price go up? Why did you get the flu? Causal questions like these arise on a regular basis, but most people likely have not thought deeply about how to answer them.

This book helps you think about causality in a structured way: what is a cause, what are causes good for, and what is compelling evidence of causality? Author Samantha Kleinberg shows you how to develop a set of tools for thinking more critically about causes. You'll learn how to question claims, explain the past, make decisions based on causal information, and verify causes through further tests.

Whether it's figuring out what data you need, or understanding that the way you collect and prepare data affects the conclusions you can draw from it, *Why* will help you sharpen your causal inference skills.

> "Philosophy, economics, statistics, and logic all try to make sense of causality; Kleinberg manages to tie together these disparate approaches in a way that's straightforward and practical. As more of our lives become 'data driven,' clear thinking about inferring causality from observations will be needed for understanding policy, health, and the world around us."
>
> —**Chris Wiggins, PhD**, Chief Data Scientist at the *New York Times* and Associate Professor at Columbia University

Samantha Kleinberg is an Assistant Professor of Computer Science at Stevens Institute of Technology. She received her PhD in computer science from New York University.



DATA / DATA SCIENCE

US $29.99 **CAN $34.99**

ISBN: 978-1-491-94964-1

Twitter: @oreillymedia
facebook.com/oreilly

O'Reilly ebooks.

Your bookshelf on your devices.

PDF
Mobi
ePub
DAISY

When you buy an ebook through oreilly.com you get lifetime access to the book, and whenever possible we provide it to you in four DRM-free file formats—PDF, .epub, Kindle-compatible .mobi, and DAISY—that you can use on the devices of your choice. Our ebook files are fully searchable, and you can cut-and-paste and print them. We also alert you when we've updated the files with corrections and additions.

Learn more at ebooks.oreilly.com

You can also purchase O'Reilly ebooks through the iBookstore, the Android Marketplace, and Amazon.com.

O'REILLY®

©2015 O'Reilly Media, Inc. The O'Reilly logo is a registered trademark of O'Reilly Media, Inc. 15055

Why: A Guide to Finding and Using Causes
by Samantha Kleinberg

Copyright © 2016 Samantha Kleinberg. All rights reserved.

Printed in the United States of America.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

O'Reilly books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (*http://safaribooksonline.com*). For more information, contact our corporate/institutional sales department: 800-998-9938 or *corporate@oreilly.com*.

Acquisitions Editor: Mike Loukides
Editor: Marie Beaugureau
Production Editor: Matthew Hacker
Copyeditor: Phil Dangler
Proofreader: Rachel Head
Indexer: Judith McConville
Interior Designer: David Futato
Cover Designer: Anton Khodakovsky
Illustrator: Samantha Kleinberg

December 2015: First Edition

Revision History for the First Edition
2015-11-13 First Release

See *http://oreilly.com/catalog/errata.csp?isbn=9781491949641* for release details.

While the publisher and the author have used good faith efforts to ensure that the information and instructions contained in this work are accurate, the publisher and the author disclaim all responsibility for errors or omissions, including without limitation responsibility for damages resulting from the use of or reliance on this work. Use of the information and instructions contained in this work is at your own risk. If any code samples or other technology this work contains or describes is subject to open source licenses or the intellectual property rights of others, it is your responsibility to ensure that your use thereof complies with such licenses and/or rights.

978-1-4919-4964-1

[LSI]

Contents

| Preface vii

1 | Beginnings 1
Where do our concepts of causality and methods for finding it come from?

2 | Psychology 17
How do people learn about causes?

3 | Correlation 35
Why are so many causal statements wrong?

4 | Time 57
How does time affect our ability to perceive and reason with causality?

5 | Observation 77
How can we learn about causes just by watching how things work?

6 | Computation 103
How can the process of finding causes be automated?

7 | Experimentation 133
How can we find causes by intervening on people and systems?

8 | Explanation 155
What does it mean to say that this caused that?

9 | Action 177
How do we get from causes to decisions?

10 | Onward 195
Why causality now?

| Notes 209

| Bibliography 233

| Index 267

Beginnings

Where do our concepts of causality and methods for finding it come from?

In 1999, a British solicitor named Sally Clark was convicted of murdering her two children. A few years earlier, in December 1996, her first son died suddenly at 11 weeks of age. At the time this was ruled as a death by natural causes, but just over a year after the first child's death, Clark's second son died at 8 weeks of age. In both cases the children seemed otherwise healthy, so their sudden deaths raised suspicions.

There were many commonalities in the circumstances: the children died at similar ages, Clark was the one who found them dead, she was home alone with the children, and both had injuries according to the post-mortem examination. The first child's injuries were initially explained as being due to resuscitation attempts, but after the second death the injuries were reexamined and now considered suspicious. Four weeks after the second death, both parents were arrested and Clark was later charged with and convicted of murder.

What are the odds of two infants in one family both dying from sudden infant death syndrome (SIDS)? According to prosecutors in the UK, this event is so unlikely that two such deaths would have to be the result of murder. This argument—that one cause is so improbable that another must have occurred—led to this now-famous wrongful conviction. It is also a key example of the consequences of bad statistics and ignoring causality.

The primary reason this case has become well known among statisticians and researchers in causality is that the prosecution created an argument centered on, essentially, the defense's explanation being too unlikely to be true. The prosecution called an expert witness, Dr. Roy Meadow, who testified that the probability of two SIDS deaths (cot death, in the UK) in one family is 1 in 73 million. Prosecutors then argued that because this probability is so low, the deaths could not have been due to natural causes and must instead have been the result of murder.

However, this statistic is completely wrong, and even if it were right, it should not have been used the way it was used. Meadow took a report estimating the chance of SIDS as 1 in 8,543 and then said the probability of two deaths is 1 in 8,543×8,543—approximately 73 million.[1] The reason this calculation is incorrect is that it assumes the events are independent. When you flip a coin, whether it comes up heads has no bearing on whether the next flip will be heads or tails. Since the probability of each is always one half, it is mathematically correct to multiply them together if we want to know the probability of two heads in a row. This is what Meadow did.

The cause of SIDS is not known for sure, but risk factors include the child's environment, such as family smoking and alcohol use. This means that given one SIDS death in a family, another is much more likely than 1 in 8,543 because the children will share the same general environment and genetics. That is, the first death gives us information about the probability of the second. This case, then, is more like the odds of an actor winning a second Academy Award. Awards are not randomly given out; rather, the same qualities that lead to someone winning the first one—talent, name recognition, connections—may make a second more likely.

This was the crux of the problem in Clark's case. Because the events are not independent and there may be a shared cause of both, it is inappropriate to calculate the probability with this simple multiplication. Instead, the probability of the second death needs to take into account that the first has occurred, so we would need to know the likelihood of a SIDS death in a family that has already had one such death. The probability and the way it was used were so egregiously wrong that the defense called a statistician as an expert witness during the first appeal, and the Royal Statistical Society wrote a letter expressing its concern.[2]

However, miscalculation was not the only problem with the probability. Prosecutors attempted to equate the 1/73 million figure for the probability of an event occurring (namely, two SIDS deaths) with the probability of Clark's innocence. This type of faulty reasoning, where the probability of an event is argued to be the probability of guilt or innocence, is actually known as the prosecutor's fallacy.[3]

Yet we already know that an unlikely event has happened. The odds of two SIDS deaths are small, but the odds of two children in one family dying in infancy are also quite small. One is not simply deciding whether to accept the explanation of SIDS, but rather comparing it against an alternative explanation. It would be better, then, to compare the probability of two children in the same family being murdered (the prosecution's hypothesis) to that of two children in the same family being affected by SIDS, given what we know about the case.

The probability of two children in one family dying from SIDS is not the same as the probability of these particular children being affected. We have other facts about the case, including physical evidence, whether there was a motive for murder, and so on. These would have to be used in conjunction with the probabilistic evidence (e.g., the likelihood of murder if someone has no motive, opportunity, or weapon would surely be lower than the overall rate).[4]

Finally, any low-probability event will eventually occur given enough trials. The incorrectly low probability in Clark's case (1 in 73 million) is still more than three times that of winning the Mega Millions lottery (1 in 258 million). The odds that you in particular will win such a lottery game are low, but the odds that someone somewhere will win? Those are quite good. This means that using only probabilities to determine guilt or innocence would guarantee at least some wrongful convictions. This is because, while it is unlikely for an individual to experience these events, given the millions of families with two children worldwide, the event will happen somewhere.

Clark's conviction was finally overturned after her second appeal in January 2003, after she'd spent three years in prison.

Why is the Sally Clark case an important example of failed causal thinking? While there were many errors in how the probability was calculated, the fundamental problem was trying to use the probability of an event occurring to support a particular causal conclusion. When trying to convince someone else of a causal explanation, have you ever said "it's just too much of a coincidence" or "what are the odds?" Even though this type of reasoning crops up often—a new employee starts at your company and on the same day your stapler disappears; a psychic knows your favorite female relative's name starts with an "M"; two key witnesses remember the suspect wearing a red flannel shirt—saying something is so unlikely to happen by chance that the only reasonable explanation is a causal connection is simply incorrect. As we've seen, the probability of an unlikely event happening to an individual may be low, but the probability of it happening somewhere is not. Getting causal explanations wrong can also have severe consequences beyond wrongful convictions, such as leading to wasted time and effort exploring a drug that will never work, or yielding ineffective and costly public policies.

This book is about doing better. Rigorous causal thinking means interrogating one's assumptions, weighing evidence, investigating alternate explanations, and identifying those times when we simply cannot know why something happened. Sometimes there is just not enough information or information of the right type to judge, but being able to know and communicate that is important. At a minimum, I hope you'll become more skeptical about the causal claims you hear (we'll discuss what questions one can ask to evaluate these claims as well as red flags to watch out for), but we'll also tackle how to find causes in the first place, develop compelling evidence of causality, and use causes to guide future actions.

What is a cause?

Take a moment and try to come up with a definition of "cause."

If you are like the students in my causal inference class, you probably got halfway through your definition before you started interrupting yourself with possible objections. Perhaps you qualified your statement with phrases like "well, most of the time," or "but not in every case," or "only if…" But your answer likely included some features like a cause *bringing about* an effect, making an effect *more likely*, having the *capability* to produce an effect, or being *responsible* for an effect. There's a general idea of something being *made* to happen that otherwise wouldn't have occurred.

While it won't be correct in all cases, in this book "cause" generally means something: that makes an effect more likely, without which an effect would or could not occur, or that is capable of producing an effect under the right circumstances.

One of the earliest definitions of causes came from Aristotle, who formulated the problem as trying to answer "why" questions.[5] So if we ask why something is the case, someone might explain how the phenomenon is produced (heating water creates vapor), what it is made from (hydrogen and oxygen bond to form water), what form it takes (the essence of a chair is something raised off the ground that has a back and is for one person to sit on), or why it is done (the purpose of a vaccine is preventing disease). Yet when we seek causes, what we often want to know is why one thing happened instead of another.

While there were other intermediate milestones after Aristotle (such as Aquinas's work in the 13th century), the next major leap forward was during the scientific revolution, toward the end of the Renaissance. This period saw major advances from Galileo, Newton, Locke, and others, but it was David Hume's work in the 18th century that became fundamental to all of our current thinking on causality and our methods for finding it.[6] That's not to say Hume got everything right (or that everyone agrees with him—or even agrees on what he believed), but he reframed the question in a critical way.

Instead of asking only what makes something a cause, Hume separated this into two questions: *what is a cause*? and *how can we find causes*? More importantly, though, instead of seeking some special feature that distinguishes causes from non-causes, Hume distilled the relationship down to, essentially, regular occurrence. That is, we learn about causal relationships by regularly observing patterns of occurrence, and we can learn about causes only through experiencing these regular occurrences.

While a mosquito bite is a necessary precursor to malaria, the sudden uptick in ice cream vendors in the spring, on the other hand, is not necessary for the weather to get warmer. Yet through observation alone, we cannot see the difference between regular occurrence (weather/ice cream) and necessity (mosquito/malaria). Only by seeing a counterexample, such as an instance of warm weather not preceded by a surge in ice cream stands, can we learn that the vendors are not necessary to the change in temperature.

It's taken for granted here that the cause happens before, rather than after or at the same time as the effect. We'll discuss this more in Chapter 4 with examples of simultaneous causation from physics, but it's important to note other ways a cause may not seem to happen before its effect. Specifically, our observation of the timing of events may not be faithful to the actual timing or the relationship itself. When a gun fires, a flash and then a loud noise follow. We may be led to believe, then, that the flash causes the noise since it always precedes the sound, but of course the gun being fired causes both of these events. Only by appealing to the common cause of the two events can we understand this regularity.

In other cases we may not be able to observe events at the time they actually occur, so they may appear to be simultaneous, even if one actually takes place before the other. This happens often in data from medical records, where a patient may present with a list of symptoms that are then noted alongside their medications. It may seem that the symptoms, diagnoses, and their prescriptions are happening simultaneously (as they're recorded during one visit), even if the medication was actually taken before symptoms developed (leading to the visit). Timings may also be incorrect due to data being collected not at the time of the event, but rather from recollection after the fact. If I ask when your last headache was, unless you made notes or it was very recent and fresh in your mind, the timing you report may deviate from the true timing, and will likely be less reliable as time passes after the event.[7] Yet to determine whether a medication is actually causing side effects, the ordering of events is one of the most critical pieces of information.

Finally, Hume requires that not only is the cause earlier than the effect, but that cause and effect should be nearby (contiguous) in both time and space. It would be difficult to learn about a causal relationship with a long delay or with the cause far removed from the effect, as many other factors may intervene in between the two events and have an impact on the outcome. Imagine a friend borrows your espresso machine, and two months after she returns it you find that it's broken. It would be much harder to pin the damage on your friend than it would be if she'd returned the machine broken (in fact, psychological experiments show exactly this phenomenon when people are asked to infer causal relationships from observations with varying time delays[8]). Similarly, if a person is standing a few feet away from a bookcase when a book falls off the shelf, it seems much less likely that he was the cause of the book falling than a person standing much closer to the shelf. On the other hand, when a pool cue hits a billiard ball, the ball immediately begins to travel across the table, making this connection much easier to discern.

The challenge to this proximity requirement is that some causal relationships do not fit this pattern, limiting the cases the theory applies to and our ability to make inferences. For example, there is no contiguity in the sense Hume stipulates when the absence of a factor causes an effect, such as lack of vitamin C causing scurvy. If we allow that a psychological state (such as a belief or intention) can be a cause, then we have another case of a true causal relationship with no physical chain between cause and effect. A student may do homework because he wants to earn an A in a class. Yet the cause of doing homework is the desire for a good grade, and there's not a physical connection between this desire and taking the action. Some processes may also occur over very long timescales, such as the delay between an environmental exposure and later health problems.

Even if there's a chain of intermediate contiguous events, we do not actually observe this chain.[9]

In Hume's view, repeatedly seeing someone pushing a buzzer and then hearing a noise (constant conjunction) is what leads you to infer that pushing the buzzer results in the noise. You make the inference because you see the person's finger make contact (spatial contiguity) with the button, this contact happens before the noise (temporal priority), and the noise results nearly immediately after (temporal contiguity). On the other hand, if there was a long delay, or the events happened at the same time, or the noise didn't always result, Hume's view is that you could not make this inference. We also could not say that pushing the button is essential to the noise, only that we regularly observe this sequence of events. There's more to the story, as we'll discuss in Chapter 5, but the basic idea here is to distinguish 1) between a cause being necessary for its effect to occur and merely seeing that a cause is regularly followed by its effect, and 2) between what the underlying relationship is and what we can learn from observation.

Note that not everyone agreed with Hume. Kant, in particular, famously disagreed with the very idea of reducing causality to regularities, arguing that necessity is the essential feature of a causal relationship and because we can never infer necessity empirically, causes cannot be induced from observations. Rather, he believed, we use a priori knowledge to interpret observations causally.[10]

* * *

While most definitions of causality are based on Hume's work, none of the ones we can come up with cover all possible cases and each one has counterexamples another does not. For instance, a medication may lead to side effects in only a small fraction of users (so we can't assume that a cause will always produce an effect), and seat belts normally prevent death but can cause it in some car accidents (so we need to allow for factors that can have mixed producer/preventer roles depending on context).

The question often boils down to whether we should see causes as a fundamental building block or force of the world (that can't be further reduced to any other laws), or if this structure is something we impose. As with nearly every facet of causality, there is disagreement on this point (and even disagreement about whether particular theories are compatible with this notion, which is called causal realism). Some have felt that causes are so hard to find as for the search to be hopeless and, further, that once we have some physical laws, those are more useful than causes anyway. That is, "causes" may be a mere shorthand for things like triggers, pushes, repels, prevents, and so on, rather than a fundamental notion.[11]

It is somewhat surprising, given how central the idea of causality is to our daily lives, but there is simply no unified philosophical theory of what causes are, and no single foolproof computational method for finding them with absolute certainty. What makes this even more challenging is that, depending on one's definition of causality, different factors may be identified as causes in the same situation, and it may not be clear what the ground truth is.

Say Bob gets mugged and his attackers intend to kill him. However, in the middle of the robbery Bob has a heart attack and subsequently dies. One could blame the mechanism (heart attack), and trace the heart attack back to its roots in a genetic predisposition that leads to heart attack deaths with high probability, or blame the mugging, as without it the heart attack would not have occurred. Each approach leads to a different explanation, and it is not immediately obvious whether one is preferable or if these are simply different ways of looking at a situation. Further, the very idea of trying to isolate a single cause may be misguided. Perhaps the heart attack and robbery together contributed to the death and their impacts cannot be separated. This assessment of relative responsibility and blame will come up again in Chapters 8 and 9, when we want to find causes of specific events (why did a particular war happen?) and figure out whether policies are effective (did banning smoking in bars improve population health in New York City?).

Despite the challenges in defining and finding causes, this problem is not impossible or hopeless. While the answers are not nearly as clear-cut as one might hope (there will never be a black box where you put in data and output causes with no errors and absolute certainty), a large part of our work is just figuring out which approach to use and when. The plurality of viewpoints has led to a number of more or less valid approaches that simply work differently and may be appropriate for different situations. Knowing more than one of these and how they complement one another gives more ways to assess a situation. Some may cover more cases than others (or cases that are important to you), but it's important to remember that none are flawless. Ultimately, while finding causes is difficult, a big part of the problem is insisting on finding causes with absolute certainty. If we accept that we may make some errors and instead aim to be explicit about what it is we can find and when, then we can try, over time, to expand the types of scenarios methods can handle, and will at least be able to accurately describe methods and results. This book focuses on laying out the benefits and limitations of the various approaches, rather than making methodological recommendations, since these are not absolute. Some approaches do better than others with incomplete data, while others may be preferable for situations in which the timing of events is important. As with much in causality, the answer is usually “it depends.”

Causal thinking is central to the sciences, law, medicine, and other areas (indeed, it’s hard to think of a field where there is no interest in or need for causes), but one of the downsides to this is that the methods and language used to describe causes can become overly specialized and seem domain-specific. You might not think that neuroscience and economics have much in common, or that computer science can address psychological questions, but these are just a few of the growing areas of cross-disciplinary work on causality. However, all of these share a common origin in philosophy.

How can we find causes?

Philosophers have long focused on the question of what causes actually are, though the main philosophical approaches for defining causality and computational methods for finding it from data that we use today didn’t arise until the 1970s and ’80s. While it’s not clear whether there will ever be a single theory of causality, it is important to understand the meaning of this concept that is so widely used, so we can think and communicate more clearly about it. Any advances here will also have implications for work in computer science and other areas. If causation isn’t just one thing, for example, then we’ll likely need multiple methods to find and describe it, and different types of experiments to test people’s intuitions about it.

Since Hume, the primary challenge has been: how do we distinguish between causal and non-causal patterns of occurrence? Building on Hume’s work, three main methods emerged during the 1960s and ’70s. It’s rarely the case that a single cause has the ability to act alone to produce an effect, so instead John L. Mackie developed a theory that represents sets of conditions that together produce an effect.[12] This better excludes non-causal relationships and accounts for the complexity of causes. Similarly, many causal relationships involve an element of chance, where causes may merely make their effects more likely without necessitating that they occur in every instance, leading to the probabilistic approaches of Patrick Suppes and others.[13] Hume also gave rise to the counterfactual approach, which seeks to define causes in terms of how things would have been different had the cause not occurred.[14] This is like when we say someone was responsible for winning a game, as without that person's efforts it would not have been won.

All of this work in philosophy may seem divorced from computational methods, but these different ways of thinking about causes give us multiple ways of finding evidence of causality. For computer scientists, one of the holy grails of artificial intelligence is being able to automate human reasoning. A key component of this is finding causes and using them to form explanations. This work has innumerable practical applications, from robotics (as robots need to have models of the world to plan actions and predict their effects) to advertisement (Amazon can target their recommendations better if they know what makes you hit "buy now") to medicine (alerting intensive care unit doctors to why there is a sudden change in a patient's health status). Yet to develop algorithms (sequences of steps to solve a problem), we need a precise specification of the problem. To create computer programs that can find causes, we need a working definition of what causes are.

In the 1980s, computer scientists led by Judea Pearl showed that philosophical theories that define causal relationships in terms of probabilities can be represented with graphs, which allow both a visual representation of causal relationships and a way to encode the mathematical relationships between variables. More importantly, they introduced methods for building these graphs based on prior knowledge and methods for finding them from data.[15] This opened the door to many new questions. Can we find relationships when there's a variable delay between cause and effect? If the relationships themselves change over time, what can we learn? Computer scientists have also developed methods for automating the process of finding explanations and methods for testing explanations against a model. Despite many advances over the past few decades, many challenges remain—particularly as our lives become more data-driven. Instead of carefully curated datasets collected solely for research, we now have a plethora of massive, uncertain, observational data. Imagine the seemingly simple problem of trying to learn about people's relationships from Facebook data. The first challenge is that not everyone uses Facebook, so you can study only a subset of the population, which may not be representative of the population as a whole or the particular subpopulation you're interested in. Then, not everyone uses Facebook in the same way. Some people never indicate their relationship status, some people may lie, and others may not keep their profiles up-to-date.

Key open problems in causal inference include finding causes from data that are uncertain or have missing variables and observations (if we don't observe smoking, will we erroneously find other factors to cause lung cancer?), finding complex relationships (what happens when a sequence of events is required to produce an effect?), and finding causes and effects of infrequent events (what caused the stock market flash crash of 2010?).

Interestingly, massive data such as from electronic health records are bringing epidemiology and computational work on health together to understand factors that affect population health. The availability of long-term data on the health of large populations—their diagnoses, symptoms, medication usage, environmental exposures, and so on—is of enormous benefit to research trying to understand factors affecting health and then using this understanding to guide public health interventions. The challenges here are both in study design (traditionally a focus of epidemiology) and in efficient and reliable inference from large datasets (a primary focus of computer science). Given its goals, epidemiology has had a long history of developing methods for finding causes, from James Lind randomizing sailors to find causes of scurvy,[16] to John Snow finding contaminated water pumps as a cause of cholera in London,[17] to the development of Koch's postulates that established a causal link between bacteria and tuberculosis,[18] to Austin Bradford Hill's linking smoking to lung cancer and creating guidelines for evaluating causal claims.[19]

Similarly, medical research is now more data-driven than ever. Hospitals as well as individual practices and providers are transitioning patient records from paper charts to electronic formats, and must meet certain meaningful use criteria (such as using the data to help doctors make decisions) to qualify for incentives that offset the cost of this transition. Yet many of the tasks to achieve these criteria involve analyzing large, complex data, requiring computational methods.

Neuroscientists can collect massive amounts of data on brain activity through EEG and fMRI recordings, and are using methods from both computer science and economics to analyze these. Data from EEG records are essentially quantitative, numerical recordings of brain activity, which is structurally not that different from stock market data, where we may have prices of stocks and volume of trades over time. Clive Granger developed a theory of causality in economic time series (and later won a Nobel Prize for this work), but the method is not specific to economics and has been applied to other biological data, such as gene expression arrays (which measure how active genes are over time).[20]

A key challenge in economics is determining whether a policy, if enacted, will achieve a goal. This is very similar to concerns in public health, such as trying to determine whether reducing the size of sodas sold will reduce obesity. Yet this problem is one of the most difficult we face. In many cases, enacting the policy itself changes the system. As we will see in Chapter 9, the hasty way a class size reduction program was implemented in California led to very different results than the original class size reduction experiment in Tennessee. An intervention may have a positive effect if everything stays the same, but the new policy can also change people's behavior. If seat belt laws lead to more reckless driving, it becomes more challenging to figure out the impact of the laws and determine whether to overturn them or enact further legislation if the death rate actually goes up.

Finally, for psychologists, understanding causal reasoning—how it develops, what differences there are between animals and humans, when it goes wrong—is one of the keys to understanding human behavior. Economists too want to understand why people behave as they do, particularly when it comes to their decision-making processes. Most recently, psychologists and philosophers have worked together using experimental methods to survey people's intuitions about causality (this falls under the umbrella of what's been called experimental philosophy, or X-Phi[21]). One key problem is disentangling the relationship between causal and moral judgment. If someone fabricates data in a grant proposal that gets funded, and other honest and worthy scientists are not funded because there is a limited pool of money, did the cheater cause them not to be funded? We can then ask if that person is to blame and whether our opinions about the situation would change if everyone else cheats as well. Understanding how we make causal judgments is important not just to better make sense of how people think, but also for practical reasons like resolving disagreements, improving education and training,[22] and ensuring fair jury trials. As we'll see throughout this book, it is impossible to remove all sources of bias and error, but we can become better at spotting cases where these factors may intrude and considering their effects.

Why do we need causes?

Causes are difficult to define and find, so what are they good for—and why do we need them? There are three main things that either can be done only with causes, or can be done most successfully with causes: prediction, explanation, and intervention.

First, let's say we want to predict who will win a presidential election. Pundits find all sorts of patterns, such as a Republican must win Ohio to win the election, no president since FDR has been reelected when the unemployment rate is over 7.2%,[23] or only men have ever won presidential elections in the US (as of the time of writing, at least).[24] However, these are only patterns. We could have found any number of common features between a set of people who were elected, but they don't tell us *why* a candidate has won. Are people voting based on the unemployment rate, or does this simply provide indirect information about the state of the country and economy, suggesting people may seek change when unemployment is high? Even worse, if the relationships found are just a coincidence, they will eventually fail unexpectedly. It also draws from a small dataset; the US has only had 44 presidents, fewer than half of whom have been reelected.

This is the problem with black boxes, where we put some data in and get some predictions out with no explanation for the predictions or why they should be believed. If we don't know why these predictions work (why does winning a particular state lead to winning the election?), we can never anticipate their failure. On the other hand, if we know that, say, Ohio "decides" an election simply because its demographics are very representative of the nation as a whole and it is not consistently aligned with one political party, we can anticipate that if there is a huge change in the composition of Ohio's population due to immigration, the reason why it used to be predictive no longer holds. We can also conduct a national poll to get a more direct and accurate measure, if the state is only an indirect indicator of national trends. In general, causes provide more robust ways of forecasting events than do correlations.

As a second example, say a particular genetic variation causes both increased exercise tolerance and increased immune response. Then we might find that increased exercise tolerance is a good indicator of someone's immune response. However, degree of exercise tolerance would be a very rough estimate, as it has many causes other than the mutation (such as congestive heart failure). Thus, using only exercise tolerance as a diagnostic may lead to many errors, incorrectly over-or underestimating risk. More importantly, knowing that the genetic variation causes both yields two ways to measure risk, and ensures we can avoid collecting redundant measurements. It would be unnecessary to test for both the gene and exercise tolerance, since the latter is just telling us about the presence of the former. Note, though, that this would not be the case if the genetic tests were highly error-prone. If that were true then exercise data might indeed provide corroborating evidence. Finally, it may be more expensive to send a patient to an exercise physiology lab than to test for a single genetic variant. Yet, we couldn't weigh the directness of a measure versus its cost (if exercise testing were much cheaper than genetic testing, we might be inclined to start there even though it's indirect) unless we know the underlying causal relationships between these factors. Thus, even if we only aim to make predictions, such as who will win an election or what a patient's risk of disease is, understanding why factors are predictive can improve both the accuracy and cost of decision-making.

Now say we want to know why some events are related. What's the connection between blurred vision and weight loss? Knowing only that they often appear together doesn't tell us the whole story. Only by finding that they share a cause—diabetes—can we make sense of this relationship. The need for causes in this type of explanation may seem obvious, but it is something we engage in constantly and rarely examine in depth.

You may read a study that says consumption of red meat is linked to a higher mortality rate, but without knowing why that is, you can't actually use this information. Perhaps meat eaters are more likely to drink alcohol or avoid exercise, which are themselves factors that affect mortality. Similarly, even if the increase in mortality is not due to correlation with other risk factors, but has something to do with the meat, there would be very different ways to reduce this hazard depending on whether the increase in mortality is due to barbecue accidents or due to consumption of the meat itself (e.g., cooking meat in different ways versus becoming vegetarian). What we really want to know is not just that red meat is linked with death, but that it is in fact causing it. I highlight this type of statement because nearly every week the science sections of newspapers contain claims involving diet and health (eggs causing or preventing various ailments, coffee increasing or decreasing risk of death). Some studies may provide evidence beyond just correlation in some populations, but all merit skepticism and a critical investigation of their details, particularly when trying to use them to inform policies and actions (this is the focus of Chapter 9).

In other cases, we aim to explain single events. Why were you late to work? Why did someone become ill? Why did one nation invade another? In these cases, we want to know who or what is responsible for something occurring. Knowing that traffic accompanies lateness, people develop various illnesses as they age, and many wars are based on ideological differences doesn't tell us why these specific events happened. It may be that you were late because your car broke down, that Jane became ill due to food poisoning, and that a particular war was over territory or resources.

Getting to the root of why some particular event happened is important for future policy making (Jane may now avoid the restaurant that made her ill, but not the particular food she ate if the poisoning was due to poor hygiene at the restaurant) and assessing responsibility (who should Jane blame for her illness?), yet it can also be critical for reacting to an event. A number of diseases and medications prescribed for them can cause the same symptoms. Say that chronic kidney disease can lead to renal failure, but a medication prescribed for it can also, in rare cases, cause the same kidney damage. If a clinician sees a patient with the disease taking this medication, she needs to know specifically whether the disease is being caused by the medication in this patient to determine an appropriate treatment regimen. Knowing it is generally possible for kidney disease to occur as a result of taking medication doesn't tell her whether that's true for this patient, yet that's precisely the information required to make a decision about whether to discontinue the medication.

Potentially the most important use of causal knowledge is for intervention. We don't just want to learn why things happen; we want to use this information to prevent or produce outcomes. You may want to know how to modify your diet to improve your health. Should you take vitamins? Become vegetarian? Cut out carbohydrates? If these interventions are not capable of producing the outcome you want, you can avoid making expensive or time-consuming changes. Similarly, we must consider degrees. Maybe you hear that a diet plan has a 100% success rate for weight loss. Before making any decisions based on this claim, it helps to know how much weight was lost, how this differed between individuals, and how the results compare to other diets (simply being cognizant of food choices may lead to weight loss). We both want to evaluate whether interventions already taken were effective (did posting calorie counts in New York City improve population health?) and predict the effects of potential future interventions (what will happen if sodium is lowered in fast food?).

Governments need to determine how their policies will affect the population, and similarly must develop plans to bring about the changes they desire. Say researchers find that a diet high in sodium is linked to obesity. As a result, lawmakers decide to pass legislation aimed at reducing sodium in restaurants and packaged foods. This policy will be completely ineffective if the only reason sodium and obesity are linked is because high-calorie fast food is the true cause and happens to be high in sodium. The fast food will still be consumed and should have been targeted directly to begin with. We must be sure that interventions target causes that can actually affect outcomes. If we intervene only on something correlated with the effect (for instance, banning matches to reduce lung cancer deaths due to smoking), then the interventions will be ineffective.

As we'll discuss later, it gets more complicated when interventions have side effects. So, we need to know not only the causes of an outcome, but also the effects of the outcome as well. For example, increasing physical activity leads to weight loss, but what's called the compensation effect can lead people to consume more calories than they're burning (and thus not only not lose weight, but actually gain weight). Rather than finding isolated links between individual variables, we need to develop a broader picture of the interconnected relationships.

What next?

Why are people prone to seeing correlations where none exist? How do juries assess the causes for crimes? How can we design experiments to figure out which medication an individual should take? As more of our world becomes driven by data and algorithms, knowing how to think causally is not going to be optional. This skill is required for both extracting useful information from data and navigating everyday decision-making. Even if you do not do research or analyze data at work, the potential uses of causal inference may affect what data you share about yourself and with whom.

To reliably find and use causes, we need to understand the psychology of causation (how we perceive and reason about causes), how to evaluate evidence (whether from observations or experiments), and how to apply that knowledge to make decisions. In particular, we will examine how the data we gather—and how we manipulate these data—affects the conclusions that can be drawn from it. In this book we explore the types of arguments that can be assembled for and against causality (playing both defense and prosecution), how to go beyond circumstantial evidence using what we learn about the signs of causality, and how to reliably find and understand these signs.

Want to read more?

You can buy this book at oreilly.com in print and ebook format.

Buy 2 books, get the 3rd FREE!

Use discount code OPC10

All orders over $29.95 qualify for **free shipping** within the US.

It's also available at your favorite book retailer, including the iBookstore, the Android Marketplace, and Amazon.com.

O'REILLY®

©2015 O'Reilly Media, Inc. The O'Reilly logo is a registered trademark of O'Reilly Media, Inc. 15055