

# Chapter 1

## Causes and Causation

### How does a cause work?

You might have expected this chapter to start with a definition of “a cause” or a definition of “to cause”; but is it essential? I think it’s not. You surely have intuitive understanding of cause-and-effect because we are all born equipped with that understanding—it is imprinted in our nature. Even a toddler who can barely speak and has never pronounced the word “cause” knows after short experience that certain actions cause certain events: pressing the right button turns the lights on or off, most of the time. Yelling “mom” often brings mom (or dad) rushing to the kids’ room. Touching the hot stove causes a burn.

But perhaps our sense of cause-and-effect is just an illusion, a misleading intuition. What if we are living in our dreams and there are no causes in sight? That’s not a joke I have made up but a possible viewpoint of a skeptic who may boldly assert that causes exist in our minds. Since the action of causation cannot be observed, it may all be perception, or speculation, or even plain imagination.

Two responses may be given to the skeptic: one funny and another serious. The funny response invites him to kick a wall as hard as possible with his bare foot. Then, we ask him whether the cause of the pain was imaginary or real. (Not to worry: “his” and “her” will take turns.) On a serious note, however, the answer is this. We pursue research of causes as if causes do operate in the objective world, and if they don’t, we are moving on a wrong path—a risk that we take in scientific inquiry. Of course, history provides some reassurance that we are not moving on a totally wrong path because causal inquiry has taken us a long way from the caves in which our ancestors lived.

Definitions and skepticism aside, we cannot ignore the question of how causes work. It turns out that key statistical methods and interpretation of research results thrive on the answer we give to this question. Different answers prescribe different methods of analysis, and sometimes a particular answer will shake the foundation of a whole set of research methods.

“How does a cause work?” is a question with no certain answer. Nor can we hope to ever know the answer. Nonetheless, two opposing *models* have been proposed: one is called determinism; the other, indeterminism.

### Deterministic causation

A deterministic model of causation belongs to a time-honored view of the world, usually attributed to Laplace, a French mathematician and philosopher (1749-1827). According to this view, the world is one big causal machine in which every state to its last detail was fully determined by the previous state—all the way back to time zero. Had we just known one state in detail and every hidden causal law of Nature, we would have known every detail of every future state. For example, you could have written this book for me because you would have known the order of the letters on each page. You could also have written every draft version that showed on my computer screen.

In Laplace's words:

"We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at any given moment knew all of the forces that animate nature and the mutual positions of the beings that compose it, if this intellect were vast enough to submit the data to analysis, could condense into a single formula the movement of the greatest bodies of the universe and that of the lightest atom; for such an intellect nothing could be uncertain and the future just like the past would be present before its eyes."

From this scary view of the world (Laplace's knows-all Demon), we derive the following key tenet of deterministic causation: every event must have been caused by a *sufficient cause*, a term best explained by an example.

Consider a sad story of 16-year-old boy who drove his dad's car in Chicago on a cold January morning. The road was slippery, he was driving fast, and the car crashed into a telephone pole. What was the cause of the crash? To answer this question, we tend to run a series of "what if" questions through our minds. What if he was driving at a speed of 20 miles-per-hour? What if it didn't snow the day before? What if his parents talked to him that morning about responsible driving?

What are we doing in this mental exercise? We are trying to view an imaginary videotape of the kid's life up to the time of the crash, and then change one thing at a time to see if the imaginary end would have been different. But are we indeed looking for just one cause? For just one culprit? Don't we have a strong feeling that many single changes to the story could have prevented the crash? Perhaps a different car, or a less slippery road, or slower speed. In other words, it is possible that many causes were crucial to bring about the crash, or as one of my colleagues says: "Many things conspire to cause an event."

A complete set of causes, each of which is essential to bringing about an event, is called a *sufficient cause* of that event because sufficiency in logic means "must follow" or "must happen". Once the set is complete (and it may take a lifetime to reach that moment), the event is inevitable; it will happen. Each piece of the set—a slippery road or speeding for example—is called a *component cause* of the event. And although the number of component causes of any particular event is unknown, nothing in this model preclude  $10^{10}$  component causes of the crash in our story. Eliminate one, any one, and there is no sufficient cause—there is no crash at the end of the videotape.

We may draw an analogy between component causes and elements of an electrical circuit. We know that each element is needed to produce current and that each element is equally important regardless of the time it was added to the circuit. Elements inserted earlier are just as important as those inserted later because if an earlier connection is disconnected, no last element will fire the electrons. In our sad story of the crash, driving fast was as important as the slipperiness of the road, assuming that both count as component causes of that crash.

Unfortunately, we cannot reconstruct a causal puzzle to find a sufficient cause, because we never know which piece was essential and which was not. Consider smoking for example. There is no way to tell if smoking was a component cause of my father's lung cancer because we cannot pull out the piece called "smoking" to see whether his lung cancer would not have occurred had he not smoked. To know a cause, we need to know the answer to a "what if" question ("what if my father did not smoke?"), but a "what if" question asks about something that did not happen—about a counter-factual world.

Before turning to an indeterministic model of causation, we may ask whether there is a sufficient cause that contains just one cause. The answer is "Maybe", but watch for the following trap. It is common to hear people talking about diseases that have a

multifactorial origin, which implies that other diseases are unifactorial: resulting from one cause rather than from ten or ten million component causes. I am not sure how many diseases truly fit the unifactorial model. First, since we don't know causes with certainty, how can we attribute any effect to a single cause? Second, a so-called single cause often turns out to be just part of the disease in question. Trisomy 21, for example, is part of Down Syndrome (whereas the cause of trisomy 21 may well be multifactorial). It seems to me that if the world is a deterministic causal machine, unifactorial causation should be rare if it exists at all, and therefore, describing a disease as multifactorial sounds like truism. Don't fall in that trap.

## Indeterministic causation

Put yourself in the passenger seat of that kid's car when he is about to leave home. Now imagine that you hold in your hands a *risk-o-meter*, a device that quantifies the objective risk of a crash on a scale from 0 to 1. You watch the driver, observe the situation, and read the display of the risk-o-meter. When the engine starts but the car has not moved yet, the reading is close to zero: it is difficult to crash a car that does not move, assuming no one else is driving around. When the car is driven out of the garage, the risk number on the risk-o-meter is still very small. As the kid keeps driving the number keeps changing, sometimes going up and sometimes going down, more or less in line with your level of fear. What do you read on the risk-o-meter at the time of the crash?

The answer depends on whom you ask.

The determinist will tell you that you read "1". The crash was inevitable because a sufficient cause was completed at that very moment. All the ingredients have accumulated to cause the crash. The determinist will also tell you that your risk-o-meter had read "0" up to the moment of the crash and that in retrospect, your sense of danger at any earlier time was mistaken because no sufficient cause was completed. Since an incomplete cause does not bring about an effect, nothing in that pre-crash situation may be described as risky. There is no room to insert the term "a risky moment" in an instantaneous transition between alive and dead.

In contrast, the indeterminist will tell you that at the time of the crash your risk-o-meter could have displayed any number in the interval  $[0,1]$ , except 0 and 1. According to indeterminism the crash was chance realization (or more precisely, probabilistic realization) of causal forces that could have crashed the car at any other time during the ride. Furthermore, no amount of information about component causes could have predicted the crash at the time and place it happened, and in fact, even if nothing had changed in the story the crash might have not happened. The indeterminist will also tell you that your changing level of fear during the ride corresponded, more or less, to objective changes in the situation, to changes in causal forces toward a crash.

The indeterministic viewpoint is best illustrated by the videotape images we used earlier. If you played the same videotape of the kid's life thousands of times—changing nothing in the story!—you would have seen various ends: rides with no crash, rides with a crash earlier, rides with a crash later, and of course, rides with the crash as it happened. Moreover, you could have computed the frequency of various ends, and those frequencies

would have corresponded to the various readings on your risk-o-meter during the ride. For example: If the risk-o-meter displayed 0.75 at the moment of the actual crash, three-quarters of the videotape replays with no crash up to that moment would have shown the actual crash; one quarter would not.

Indeterminism says that a cause generates a causal force, a *propensity* or disposition to produce its effect, regardless of whether that effect is eventually realized. A slippery road, for instance, generates a propensity toward a car crash even if no car crash happens on that road. Smoking may generate a propensity toward lung cancer in every smoker even though many smokers escape that fate. What the determinist called component causes, patiently awaiting their companions to complete a sufficient cause, is viewed by the indeterminist as the sum of causal propensities that may or may not produce an effect in any particular instance. Try to think about causal propensity as a real, pushing physical force—as real as the pushing of the wind on your face.

Do I hear you saying that we cannot observe a causal propensity directly and therefore the previous paragraph reads like science fiction? True, we cannot observe that force directly, but we can't directly observe any causation. All that we can observe in the world is the *consequences* of these forces. And just as we observe the consequences of gravity (my computer is not floating in the air), we sometimes observe the consequences of causal propensities—that is, their effects. You have come to accept gravitation as real and may doubt the realness of an indeterministic causal force only because one is an old-timer and the other is new; the idea of indeterminism has not penetrated yet into high-school textbooks.

If causes indeed generate forces to bring about an effect, we may assume that these forces vary in strength, a concept that found no place in a deterministic model of causation. (Recall that determinism assigned equal importance to each component cause because any missing piece precluded the completion of a sufficient cause.) Indeterminism fully embraces the idea that causes could be ranked by order of importance, corresponding to the magnitude of their causal propensities. Perhaps a slippery road has greater propensity to cause a crash than has driving fast. Or, relative to active smoking, second-hand smoking may have a weaker propensity to cause lung cancer.

Indeterminism is not easy to digest. It seems to me that we are primarily programmed to think deterministically—we want to find the final explanation for what has happened—so it is difficult to accept that chance may lie at the end of complete knowledge. Albert Einstein, for example, refused to accept this idea most of his life, having said that "*He* [God] does not play dice" and thereby having denied the possibility of an indeterministic Nature. Indeterminism robs us from our naïve hope to predict the future precisely and some day, perhaps, to have full control over what will happen.

If you think a little longer, however, you may come to the opposite conclusion. Determinism tells us that the future is fully determined by the past, so any sense of control we have must be illusionary. Indeterminism, conversely, leaves the future open, allowing us to plan ahead and sometimes change causal forces to our advantage, albeit with no guaranteed outcome. Of the two models, indeterminism seems more optimistic to me. It also seems more favorable to the idea of free will, but not everyone agrees.

Is it possible to combine the two models? I think the answer is "No". I had thought that indeterminism entails deterministic events as a special case until someone referred me to an opposing view. I have also seen an attempt to incorporate chance (that is, indeterminism) into a model of deterministic causation but did not find that proposal convincing. Philosophical debates rarely reach closure, if ever.

## Does the world appear deterministic or indeterministic?

It depends, again, on whom you ask. If you happen to ask a scientist who conducts experiments in quantum mechanics, she will say that the world looks indeterministic at its very core. And if the invisible building blocks are indeterministic, it makes sense that the whole structure plays by the same rule. On the other hand, if you ask a computer scientist with a binary world in mind (on or off), chances are he will tell you that the world is fully deterministic and he may even belittle the whole idea of indeterminism. When I look at data from epidemiological studies, I see probabilistic regularities that are difficult to explain in a deterministic world and I end up wondering whether the world is truly indeterministic or just pretends to look so through data. Each time you say to yourself “It would not have happened if that hadn’t happened,” you are expressing a deterministic view. And each time you say to yourself “A was a more important cause of Z than B was”, you are speaking the language of indeterminism. In our lives, we seem to switch easily from one model to another without worrying too much about coherence. In Science—you cannot.

The world appeared solidly deterministic until quantum mechanics has turned the house upside down for many scientists. At the turn of the last century, however, a group of scholars in artificial intelligence, computer science, epidemiology, and statistics reversed the trend back to determinism, perhaps because a deterministic model has opened the door to a closed system of causal inference. If the model is wrong, however, a closed system of inference leads nowhere. Many scientists, including me, have hard time reconciling determinism-based methods for causal inference with our views about scientific inquiry. You will see why in the next chapters.

## A sense of conviction

To conclude this chapter, let’s consider a side question that I found interesting and may be of interest to you. If we can’t know causes with certainty, why are we often convinced that we know what has caused an event? For example, when I fell on my ice-coated driveway (back in Minnesota...), I was *sure* that the slippery driveway was the cause of my falling. What determines our level of conviction about the cause of a historical event?

This is of course a question for the field of psychology, which has nothing to do with the field of epistemology (theories of knowledge), but consider the following hypothesis. Our level of conviction that A has caused B is related to the proximity of A to B in time and space and to the magnitude of the estimated effect of A on B. The closer is A to B and the stronger the estimated effect of A on B, the greater our sense of conviction that A has caused B.

The proximity part of the hypothesis is self-explanatory and possibly intuitive, but you may rightly question the second claim. After all, neither you nor I ever saw a layperson calculating the estimated effect of a slippery driveway on falling.

Indeed the layperson does not calculate with exact numbers but I believe that his subconscious mind runs the following approximation. “The chance of my falling on a dry driveway is small—I walk on the driveway daily and rarely fall. On the other hand, the chance of my falling on an icy driveway is high, maybe 100-fold or 1000-fold higher.

People fall on icy driveways all the time. Therefore, the slippery driveway must have caused my falling.”

A sense of conviction is, of course, useless in scientific inquiry. At various times people were convinced, beyond any doubt, of the truth of statements that almost nobody believes today—that the earth is flat, for example. So every time you are deeply convinced that you know a cause (or any other scientific truth), pause for a second and recall convictions of the past: the sun circles the earth; an atom cannot be divided; Newtonian mechanics is the ultimate truth in physics. Fight your sense of conviction with all your might. That’s the only way to discover falsehood and (conjecturally) move forward.

Addendum:

Much of science, if not all of it, is concerned with discovering causal connections between variables. Does  $E \rightarrow D$ ? Or more rigorously: what is the causal parameter for the effect of E on D? Name the variables E and D, and you have a scientific theory. Sounds simple? Yes, the basic question might be simple, but providing the answer quickly turns out to be a very complicated task. Have you asked a valid question? What is E exactly? Does it exist? How do I find its value? How do I estimate a causal parameter? What kinds of biases are lurking in the background? How do I try to remove them? And so on.