305

INDUCTIVE REASONING

In previous chapters we have been concerned primarily with deductive arguments that aim at validity. However, many arguments (perhaps most arguments) encountered in daily life are not intended to meet the standard of validity. They are claimed only to provide reasons (perhaps very strong reasons) for their conclusions. Such arguments are called inductive. *This chapter will discuss five kinds of inductive reasoning. First, in* inferences to the best explanation, *explanatory and predictive power are said to provide evidence for certain hypotheses. Second, in* arguments from analogy, *the fact that two things have certain features in common is taken as evidence that they have further features in common. Third, we will develop tests for* necessary *conditions and for* sufficient *conditions, which apply to causal reasoning. Fourth, we will discuss* inductive generalizations, *where a statistical claim is made about a population on the basis of features of a sample of that population. Finally, the chapter examines* statistical syllogisms. *Here a statistical claim is made about members of a population on the basis of features of the population.*

INDUCTION VERSUS DEDUCTION

The distinction between deductive arguments and inductive arguments can be drawn in a variety of ways, but the fundamental difference concerns the relationship that is claimed to hold between the premises and the conclusion for each type of argument. An argument is *deductive* insofar as it is intended or claimed to be *valid.* As we know from Chapter 2, an argument is valid if and only if it is impossible for the conclusion to be false when its premises are true. The following is a valid deductive argument:

All ravens are black.

∴ If there is a raven on top of Pikes Peak, then it is black.

Because the premise lays down a universal principle governing all ravens, if it's true, then it *must* be true of all ravens (if any) on top of Pikes Peak.

This same relationship does not hold for invalid arguments. Nonetheless, arguments that are not valid can still be deductive if they are intended or claimed to be valid. Suppose Sally argues, "All ravens are black, and some black things are awake at night, so some ravens must be awake at night."

Although this argument is invalid, it has a form that is often mistakenly believed to be valid. The fact that Sally chose to put this argument in such a form suggests that she probably intended her argument to be valid. That intention makes the argument deductive. When someone puts forward such a deductive argument, that person is committed to the claim that the argument is valid, and so may be criticized if this relationship does not hold.

In contrast, inductive arguments are not intended to be valid, so they should not be criticized for being invalid. The following is an example of an inductive argument:

All ravens that we have observed so far are black.
$\therefore$ All ravens are black.

Here we have drawn an inductive inference from the characteristics of *observed* ravens to the characteristics of *all* ravens, most of which we have not observed. Of course, the premise of this argument *could be* true, yet the conclusion turn out to be false. A raven that has not yet been observed might be albino. The obviousness of this possibility suggests that someone who gives this argument does not put it forth as valid, so it is not a deductive argument. Instead, the premise is put forth as a *reason* or *support* for the conclusion. When an argument is not claimed to be valid but is intended only to provide a reason for the conclusion, the argument is *inductive*.

306

Because inductive arguments are supposed to provide reasons, and reasons vary in strength, inductive arguments can be evaluated as *strong* or *weak*, depending on the strength of the reasons that they provide for their conclusions. If we have seen only ten ravens, and all of them were in our back yard, then the above argument gives at most a very weak reason to believe that all ravens are black. However, if we have traveled around the world and seen over half the ravens that exist, then the above argument gives a strong reason to believe that all ravens are black. Inductive arguments are usually intended to provide strong support for their conclusions, in which case they can be criticized if the support they provide is not strong enough for the purposes at hand.

The most basic distinction, then, is not between two kinds of argument but instead between two standards for evaluating arguments. The deductive standard is validity. The inductive standard is strength. Arguments themselves are classified as either deductive or inductive in accordance with the standard that they are intended or claimed to meet. Validity is not claimed for inductive arguments, even when they put the conclusion beyond any reasonable doubt. Consequently, such inductive arguments should not be criticized as invalid, because that is a standard that they were never intended to meet. In contrast, inductive arguments can be criticized as weak if they do not give strong enough reasons for their conclusions.

There are several important differences between deductive and inductive standards that should be noticed. One fundamental feature of the deductive standard of validity is that adding premises to a valid argument cannot make it invalid. The definition of validity guarantees this: in a valid argument, it is not possible for the premises to be true without the conclusion being true as well. If any further premises could change this, then it would be possible for this relationship not to hold, so the argument would not be valid after all. Furthermore, if an argument is not only valid but sound, its soundness cannot be removed by adding further premises as long as they are true. Additional information might, of course, lead us to question the truth of one of the premises that we previously accepted, and we might reconsider whether the argument that we previously considered sound really is sound; but that is another matter.

The situation is strikingly different when we deal with inductive arguments. To cite a famous example, before the time of Captain Cook's voyage to Australia, Europeans had observed a great many swans, and every one of them was white. Thus, up to that time Europeans had very strong inductive evidence to support the claim that all swans are white. Then Captain Cook discovered black swans in Australia. What happens if we add this new piece of information to the premises of the original inductive argument? Provided that we accept Cook's report, we now produce a sound *deductive* argument in behalf of the claim that *not* all swans are white; for, if some swans are black, then not all of them are white. This, then, is a feature of every inductive argument: no matter how strong an inductive argument is, the possibility remains open that further information can undercut, perhaps completely, the support that the premises give to the conclusion. Because all inductive arguments are able to be defeated in this way, they are described as *defeasible*. Valid deductive arguments do not face a similar peril, so they are *indefeasible*.[1]

A second important difference between inductive and deductive standards is that inductive strength comes in degrees, but deductive validity does not. An argument is either valid or invalid. There is no question of how much validity an argument has. In contrast, inductive arguments can be more or less strong. The more and more varied ravens or swans we observe, the stronger the inductive arguments above. Some inductive arguments are extremely strong and put their conclusions beyond any reasonable doubt. Other inductive arguments are much weaker, even though they still have some force.

Because of the necessary relationship between the premises and the conclusion of a valid deductive argument, it is often said that the premises of valid deductive arguments (if true) provide *conclusive* support for their conclusions, whereas true premises of strong inductive arguments provide only *partial* support for their conclusions. There is something to this: because the

[1]This difference is sometimes described by saying that deductive standards are *monotonic*, whereas inductive standards are *nonmonotonic*. These labels come from an analogy with mathematical curves: a (positive) monotonic curve rises, perhaps flattens out, but never changes direction. A nonmonotonic curve can change direction, for example, rise, then fall. Similarly, the strength of an inductive argument can change up or down as further information is added, but the validity of a deductive argument is never lost by adding more premises.

premises of a valid deductive argument necessitate the truth of the conclusion, if those premises are known to be true, then they do supply conclusive reasons for the conclusion. The same cannot be said for inductive arguments.

However, it would be altogether misleading to conclude from this that inductive arguments are inherently inferior to deductive arguments in supplying a justification or ground for a conclusion. In the first place, inductive arguments often place matters beyond any reasonable doubt. It is possible that the next pot of water will not boil at any temperature, however high, but this is not something we worry about. We do not take precautions against it, and we shouldn't.

More important, deductive arguments normally enjoy no advantages over their inductive counterparts. We can see this by comparing the two following arguments:

DEDUCTIVE	INDUCTIVE
All ravens are black.	All observed ravens are black.
∴ If there is a raven on top of Pikes Peak, it is black.	∴ If there is a raven on top of Pikes Peak, it is black.

Of course, it is true for the deductive argument (and not true for the inductive argument) that if the premise is true, then the conclusion must be true. This may seem to give an advantage to the deductive argument over the inductive argument. However, before we can decide how much support a deductive argument gives its conclusion, we must ask if its premises are, after all, true. That is not something we can just take for granted. If we examine the premises of these two arguments, we see that it is easier to establish the truth of the premise of the inductive argument than it is to establish the truth of the premise of the deductive argument. If we have observed carefully and kept good records, then we might be fully confident that all *observed* ravens have been black. On the other hand, how can we show that *all* ravens (observed and unobserved—past, present, and future) are black? The most obvious way (though there may be other ways) would be to observe ravens to see if they are black or not. This, of course, involves producing an inductive argument (called an inductive generalization) for the premise of the deductive argument. Here our confidence in the truth of the premise of the deductive argument should be no greater than our confidence in the strength of the inference in the inductive generalization. In this case—and it is not unusual—the deductive argument provides no stronger grounds in support of its conclusion than does its inductive counterpart, because any reservations we might have about the *strength* of the inductive inference will be paralleled by doubts concerning the *truth* of the premise of the deductive argument.

We will also avoid the common mistake of saying that deductive arguments always move from the general to the particular, whereas inductive arguments always move from the particular to the general. In fact, both sorts of arguments can move in either direction. There are inductive arguments

307

intended to establish particular matters of fact, and there are deductive arguments that involve generalizations from particulars. For example, when scientists assemble empirical evidence to determine whether the extinction of the dinosaurs was caused by the impact of a meteor, their discussions are models of inductive reasoning. (See Chapter 15.) Yet they are not trying to establish a generalization or a scientific law. Instead, they are trying to determine whether a particular event occurred some 65 million years ago. Inductive reasoning concerning particular matters of fact occurs constantly in everyday life as well, for example, when we check to see whether our television reception is being messed up by someone using a computer. Deductive arguments from the particular to the general also exist, though they tend to be trivial, and hence boring. Here's one: Benjamin Franklin was the first postmaster general; therefore, anyone who is identical with Benjamin Franklin was the first postmaster general. Of course, many deductive arguments do move from the general to the particular, and many inductive arguments do move from particular premises to a general conclusion. It is important to remember, however, that this is not the *definitive* difference between these two kinds of arguments.

Indeed, the first kind of inductive reasoning that we will discuss often leads to conclusions about particular matters of fact. This form of argument is called *inference to the best explanation*. Inferences to the best explanation are extremely common in everyday life. We will argue that another common kind of inductive inference, which is called *argument from analogy*, is best understood as an incomplete inference to the best explanation. In our view, both of these forms of inductive argument are based on general principles. So, next, we will investigate methods for testing generalizations, for supporting generalizations by sampling, and finally for applying generalizations back to particular cases. This survey will reveal the tremendous diversity among inductive arguments.

What unites these various forms of argument and makes them all inductive is, as we said, the *claimed* relationship between the premises and the conclusion. Inductive arguments are not intended to be valid, so it is a mistake to criticize them for lacking validity. It is enough for them to meet inductive standards by being strong. Of course, this general definition does not yet tell us what inductive strength is or how to determine whether a given inductive argument is strong. That requires a longer story, which will occupy us for the rest of this chapter. This story can be told in two ways. At an informal level, we can lay down some general rules—rules of thumb—that will allow us to avoid many of the more common errors of inductive reasoning. In complicated situations, however, our commonsense principles are often inadequate and can let us down. When common sense gets out of its depth, we must turn for help to methods that are more technical, such as the procedures of mathematical statistics. In this chapter, we shall concern ourselves almost exclusively with informal procedures for the evaluation of inductive arguments, but we shall also examine some cases in which they are inadequate.

EXERCISE I

Assuming a standard context, label each of the following arguments as deductive or inductive. Explain what it is about the words or form of argument that indicates whether or not each argument is intended or claimed to be valid. If it is not clear whether the argument is inductive or deductive, say why.

(1) The sun is coming out, so the rain will probably stop soon.
(2) It's going to rain tomorrow, so it will either rain or be clear tomorrow.
(3) No woman has ever been elected president. Therefore, no woman will ever be elected president.
(4) Diet cola never keeps me awake at night. I know because I drank it just last night without any problems.
(5) The house is a mess, so Jeff must be home from college.
(6) If Harold were innocent, he would not go into hiding. Since he is hiding, he must not be innocent.
(7) Nobody in Paris seems to understand me, so either my French is rotten or Parisians are unfriendly.
(8) Because both of our yards are near rivers in Tennessee, and my yard has lots of mosquitoes, there must also be lots of mosquitoes in your yard.
(9) Most likely, her new husband speaks English with an accent, because he comes from Germany, and most Germans speak English with an accent.
(10) There is no even number smaller than two, so one is not an even number.

308

DISCUSSION QUESTIONS

(1) The following arguments are not clearly inductive and also not clearly deductive. Explain why.
   (a) All humans are mortal, and Socrates is a human, so Socrates is likely to be mortal also.
   (b) We checked every continent there is, and every raven in every continent was observed to be black, so every raven is black.
   (c) If there's radon in your basement, this monitor will go off. The monitor is going off, so there must be radon in your basement. (Said by an engineer while running the monitor in your basement.)
(2) In mathematics, proofs are sometimes employed using the method of *mathematical induction*. If you are familiar with this procedure, determine whether these proofs are deductive or inductive in character. Explain why.

INFERENCES TO THE BEST EXPLANATION

One of the most common forms of inductive argument is inference to the best explanation.[2] The general idea behind such inferences is that a hypothesis gains inductive support if, when added to our stock of previously accepted beliefs, it enables us to explain something that we observe or believe, and no competing explanation works nearly as well.

To see how inferences to the best explanation work, suppose you return to where you live and discover that the lock on your front door is broken and a number of valuable objects are missing. In all likelihood you will immediately conclude that you have been burglarized. Of course, other things *could* have produced the mess. Perhaps there was a drug bust and the authorities had the wrong address. Perhaps your friends are playing a strange joke on you. Perhaps a meteorite struck the door and then vaporized your valuables. In fact, all of these things *could* have happened (even the last), and further investigation could show that one of them did. Why, then, do we so quickly accept the burglary hypothesis without even considering these competing possibilities? The reason is that the hypothesis that your home was robbed is not highly improbable; and this assumption, taken together with other things we believe, provides the best—the strongest and the most natural—explanation of the phenomenon. The possibility that a meteorite struck your door is so wildly remote that it is not worth taking seriously. The possibility that your house was raided by mistake or that your friends are playing a strange practical joke on you is not wildly remote, but neither fits the overall facts very well. If it was a police raid, then you would expect to find a police officer there or at least a note of some kind. If it is a joke, then it is hard to see the point of it. By contrast, burglaries are not very unusual, and that hypothesis fits the facts extremely well. Logically, the situation looks like this:

(1) OBSERVATION: Your lock is broken, and valuables are missing.
(2) EXPLANATION: The hypothesis that your house has been burglarized, combined with previously accepted facts and principles, provides a suitably strong explanation of observation (1).
(3) COMPARISON: No other hypothesis provides an explanation nearly as good as that in (2).
(4) CONCLUSION: Your house was (probably) burglarized.

The explanatory power of the conclusion gives us reason to believe it because doing so increases our ability to make reliable predictions and illuminating

[2]Gilbert Harman deserves much of the credit for calling attention to the importance of inferences to the best explanation; see, for example, his *Thought* (Princeton: Princeton University Press, 1973). A similar form of argument called abduction was analyzed long ago by Charles Saunders Peirce; see, for example, his *Collected Papers of Charles Saunders Peirce* (Cambridge: Harvard University Press, 1931), 5:189. A wonderful recent discussion is Peter Lipton, *Inference to the Best Explanation* (London: Routledge, 1991).

explanations. Explanation is important because it makes sense out of things—makes them more intelligible—and we want to understand the world around us (see Chapter 3). Prediction is important because it tests our theories with new data and sometimes allows us to anticipate or even control future events. We want to know how to adjust our actions to fit what will happen. Inference to the best explanation enables us to do just that.

Here it might help to compare inferences to the best explanation with other forms of argument. In a justificatory use of argument, we cite true premises as reasons to believe the truth of the conclusion. In an explanatory use of argument, we try to make sense of something by deriving it (sometimes deductively) from premises that are themselves well understood. With an inference to the best explanation, we reason in the opposite direction from explanation: instead of deriving an observation from its explanation, we derive the explanation from the observation. That a hypothesis provides the best explanation of something whose truth is already known provides evidence for the truth of that hypothesis.

Once we grasp the notion of an inference to the best explanation, we can see this pattern of reasoning everywhere. Solutions to murder mysteries almost always have the form of an inference to the best explanation. The facts of the case are laid out and then the clever detective argues that, given these facts, only one person could possibly have committed the crime. In the story "Silver Blaze," Sherlock Holmes concludes that the trainer must have been the dastardly fellow who stole Silver Blaze, the horse favored to win the Wessex Cup, which was to be run the following day. Holmes's reasoning, as usual, was very complex, but the key part of his argument was the fact that the dog kept in the stable did not bark loudly when someone came and took away the horse.

309

> I had grasped the significance of the silence of the dog, for one true inference invariably suggests others. [I knew that] a dog was kept in the stables, and yet, though someone had been in and fetched out a horse, he had not barked enough to arouse the two lads in the loft. Obviously the midnight visitor was someone whom the dog knew well.[3]

Together with other facts, this was enough to identify the trainer, Straker, as the person who stole Silver Blaze. In this case, it is the fact that something *didn't* occur that provides the basis for an inference to the best explanation.

Of course, Holmes's inference is not absolutely airtight. It is possible that Straker is innocent and Martians with hypnotic powers over dogs committed the crime. But that only goes to show that this inference is neither valid nor deductive in our sense. It does not show anything wrong with Holmes's inference. Since his inference is inductive, it is enough for it to be strong.

Inferences to the best explanation are also defeasible. No matter how strong such an inference might be, it can always be overturned by future

[3] Sir Arthur Conan Doyle, "Silver Blaze," *The Complete Sherlock Holmes* (Garden City, New York: Doubleday, 1930), 1:349. The stories describe Holmes as a master of deduction, but his arguments are inductive as we define the terms.

experience. This can happen in a number of ways. To go back to the previous example of your ransacked house, your friends might suddenly jump out shouting, "Surprise, surprise!" Alternatively, one might think up an even better explanatory hypothesis. Maybe your roommate broke the lock and stole your valuables as part of an insurance scam. In any case, the defeasibility of inferences to the best explanation does not show that they are all bad or all on a par. If you find a broken glass on the floor, then it is a practical certainty that it broke as the result of falling. It is entirely possible that someone broke the glass with a hammer and then placed the pieces on the floor, and this hypothesis might be established by future evidence. However, unless and until such special evidence comes along, it is hard to imagine why anyone would take a hammer to the glass, so this hypothesis is very weak as an explanation. In the absence of another plausible alternative, we have adequate reason to believe that the glass fell on the floor, because that hypothesis provides the best explanation of the information that is available to us now.

We still need some way to determine which explanation is the *best*. There is, in fact, no simple rule for deciding this, but we can list some factors that go into the evaluation of an explanation.[4]

First, the hypothesis should really *explain the observations*. A good explanation makes sense out of the thing it is intended to explain. The broken lock can be explained by a burglary but not by the hypothesis that a friend came to see you (unless you have strange friends). Moreover, the hypothesis needs to explain *all* of the observations. The hypothesis of a mistaken police raid might explain the broken lock but not the missing valuables. It is also important that the explanation itself not be in need of explanation. It does not help to explain something that is obscure by citing something just as obscure. Why did the police raid your house? Because they had the wrong address. If they did not have the wrong address, then we would wonder why they suspected you. Without an explanation of *that*, the police raid hypothesis could not explain even the broken lock as well as the hypothesis of a burglary.

Second, the explanation should be *powerful*. It is a mark of excellence in an explanation that the same kind of explanation can be used successfully over a wide range of cases. Many broken locks can be explained by burglaries. Explanatory range is especially important in science. One of the main reasons why Einstein's theory of relativity replaced Newtonian physics is that Einstein could explain a wider range of phenomenon, including very small particles at very high speeds. However, explanations can go too far. Consider the hypothesis that every particle of matter has its own spirit that makes it do what it does. This might seem to explain some phenomenon that even Einstein's theory cannot explain. However, the spirit hypothesis really explains nothing, because it does not explain why any particle behaves one way as opposed to another. Either behavior is compatible with the hypothesis, so neither is explained (see also Chapter 12 on self-sealers).

[4] This discussion in many ways parallels and is indebted to the fifth chapter of W. V. Quine and J. S. Ullian, *The Web of Belief*, 2d ed. (New York: Random House, 1978).

Moreover, explanations should be *modest* in the sense that they should not claim too much, that is, any more than is needed to explain the observations. When you find your lock broken and valuables gone, you should not jump to the conclusion that there is a conspiracy against you or that gangs have taken over your neighborhood. Without further information, there is no need to specify that there was more than one burglar (or just one burglar, for that matter) in order to explain what you see. For this reason, the most modest explanation would not specify any number of burglars, so no inference to the best explanation could justify any claim about the number of burglars, at least until more evidence comes along.

Modesty is related to *simplicity*. One kind of simplicity is captured by the celebrated principle known as Occam's razor, which tells us not to multiply entities beyond necessity. Physicists, for example, should not postulate new kinds of subatomic particles or forces unless there is no other way to explain their experimental results. Similar standards apply in everyday life. We should not believe in ghosts unless they really are necessary to explain the noises in our attic or some other phenomenon. Simplicity is not always a matter of the number of entities. If you show up for class at the usual time, but nobody is there, then you might assume that everyone else independently decided to skip class that day or that they are all playing a joke on you; but it would normally be simpler to think that the class was canceled and you forgot or never heard about it. Simplicity is a mark of excellence in an explanation partly because simple explanations are easier to understand and apply, but considerations of plausibility and aesthetics are also at work in judgments of which explanation is simplest.

The tests of modesty and simplicity might seem to be in tension with the test of power. This tension can be resolved only by finding the right balance. The best explanation will not claim any more than is necessary (so it will be modest), but it will claim enough to cover a wide range of phenomena (so it will be powerful). This is tricky, but the best explanations do it.

Finally, an explanation should be *conservative*. If an explanation forces us to give up other well-established beliefs, then it should do this as little as possible. We have strong reasons to believe that cats cannot break metal locks, and this rules out the hypothesis that your front-door lock was broken by the neighbor's cat. Explanations should also not contain claims that are themselves unlikely to be true. A meteorite would be strong enough to break your lock, but it is very unlikely that a meteorite struck your lock. That makes the burglary hypothesis better, at least until we find other evidence (such as meteorite fragments) that cannot be explained except by a meteorite.

In sum, a hypothesis provides the best explanation when it is more explanatory, powerful, modest, simple, and conservative than any competing hypothesis. Each of these standards can be met to varying degrees, and they can conflict. As we saw, the desire for simplicity might have to be sacrificed to gain a more powerful explanation. Conservatism also might have to give way to explain some unexpected observations, and so on. These standards are not always easy to apply, but they can often be used to determine whether a particular explanation is better than its competitors.

Once we determine that one explanation is the best, we still cannot quite yet infer that it is true. It might turn out that the very best explanation isn't much of an explanation at all. The best explanation in a group of weak explanations does not provide much evidence for saying that the explanatory hypothesis is true. For centuries people were baffled by the floods that occurred in the Nile River each spring. The Nile, as far as anyone knew, flowed from an endless desert. Where, then, did the flood waters come from? Various wild explanations were suggested—mostly about deities of one kind or another—but none was any good. Looking for the best explanation among these weak explanations would be a waste of time. It was only after it was discovered that central Africa contains a high mountain range covered with snow in the winter that a reasonable explanation became possible. That, in fact, settled the matter. So it must be understood that the best explanation must also be a *good* explanation.

Even when an explanation is both good and best, what it explains might be illusory. Many people believe that shark cartilage prevents cancer, because the best explanation of why sharks do not get cancer lies in their cartilage. One serious problem for this inference is that sharks *do* get cancer. They even get cancer in their cartilage. So this inference to the best explanation fails.

When a particular explanation is both good and much better than any competitor, and when the explained observation is accurate, then an inference to the best explanation will have *strong* inductive support. At other times, no clear winner or even reasonable contender emerges. In such cases, an inference to the best explanation will be correspondingly *weak*.

Whether an inference to the best explanation is strong *enough* depends on the context. As contexts shift, standards of rigor can change. An inference that is strong enough to justify my belief that my spouse took our car might not be strong enough to convict our neighbor of stealing our car. Good judgment is often required to determine whether a certain degree of strength is adequate for the purposes at hand.

Context can also affect the rankings of various factors. Many explanations, for example, depend on universal premises. In such cases, compatibility with observation is usually the primary test. The universal principle should not be refuted by counterexamples (see Chapter 3). But sometimes explanatory power will take precedence: if a principle has strong explanatory power, we may accept it even in the face of clear disconfirming evidence. We do not give up good explanations lightly—nor should we. To understand why, notice that we do not test single propositions in isolation from other propositions in our system of beliefs. When faced with counterevidence to our beliefs, we often have a choice between what to give up and what to continue to hold onto. A simple example will illustrate this. Suppose that we believe the following things:

(1) Either John or Joan committed the crime.

(2) Whoever committed the crime must have had a motive for doing so.

(3) Joan had no motive to commit the crime.

From these three premises we can validly infer that John committed the crime. Suppose, however, that we discover that John could not have committed the crime. (Three bishops and a Boy Scout leader swear that he was somewhere else at the time.) Now, from the fact that John did not commit the crime, we could not immediately conclude that Joan committed it, for that would lead to an inconsistency. If she committed the crime, then, according to (3), she would have committed a motiveless crime, but that conflicts with (2), which says that motiveless crimes do not occur. So the discovery that John did not commit the crime entails that at least one of the premises in the argument must be abandoned, but it does not tell us which one or which ones.

This same phenomenon occurs when we are dealing with counterevidence to a complex system of beliefs. Counterevidence shows that there must be something wrong somewhere in the system, but it does not show exactly where the problem lies. One possibility is that the *supposed* counterevidence is itself in error. Imagine that a student carries out an experiment and gets the result that one of the fundamental laws of physics is false. This will not shake the scientific community even a little, for the best explanation of the student's result is that she messed things up. Given well-established principles, she could not have gotten the result she did if she had run the experiment correctly. Of course, if a great many reputable scientists find difficulties with a supposed law, then the situation is different. The hypothesis that all of these scientists have, like the student, simply messed up is itself highly unlikely. But it is surprising how much contrary evidence will be tolerated when dealing with a strong explanatory theory. Scientists often continue to employ a theory in the face of counterevidence. Sometimes this perpetuates errors. For years, instruments reported that the levels of ozone above Antarctica were lower than before, but scientists attributed these measurements to bad equipment, until finally they announced an ozone hole there. Still, there is often good reason to hold on to a useful theory despite counterevidence, as long as its defects do not make serious trouble, that is, give bad results in areas that count. Good judgment is required to determine when it is finally time to shift to a different explanation.

311

EXERCISE II

Which hypotheses would best explain the following phenomena? In each case, indicate how strong the explanation is. In deciding this, consider the standards of a good explanation, and compare a number of alternative explanations.

(1) Your house begins to shake so violently that pictures fall off your walls.
(2) Your key will not open the door.
(3) People start putting television cameras on your lawn, and a man with a big smile comes walking up your driveway.
(4) Virtually all of the food in markets has suddenly sold out.
(5) You put on a shirt and notice that there is no pocket on the front like there used to be.

(6) A cave is found containing the bones of both prehistoric humans and now-extinct predators.
(7) After being visited by lobbyists for cigarette producers, your senator votes in favor of tobacco price supports, although he opposed them before.
(8) Large, mysterious patterns of flattened wheat appear in the fields of Britain. (Some people attribute these patterns to visitors from another planet.)
(9) A palm reader foretells that something wonderful will happen to you soon, and it does.
(10) A neighbor sprinkles purple powder on his lawn to keep away tigers, and, sure enough, no tigers show up on his lawn.

EXERCISE III

Imagine that a contractor finished building your new home just before it fell down. You call him on the phone and ask, "Why did it fall down?" This question might be answered in any of the following ways. For each answer, say which standard of a good explanation, if any, it violates. The standards require that a good explanation be explanatory, powerful, modest, simple, and conservative. A single answer might violate more than one standard.

(1) Because its walls were not strong enough.
(2) Because an evil ghost roams the world and knocks down some buildings for no apparent reason.
(3) Because I always build shoddy buildings.
(4) Because I wanted one of my buildings to fall down, so I built it that way on purpose.
(5) Because the particular boards that I used had cracks in them, although I had no way of knowing that when I nailed them in.
(6) Because a powerful earthquake was centered on your property but did not affect anything or anybody else.

DISCUSSION QUESTIONS

(1) Put the following inference to the best explanation in standard form, and then evaluate it as carefully as you can, using the tests discussed above.

[During the Archean Era, which extended from about 3.8 to 2.5 million years before the present,] the sun's luminosity was perhaps 25% less than that of today. . . . This faint young sun has led to a paradox. There is no evidence from the scant rock record of the Archean that the planetary surface was frozen. However, if Earth had no atmosphere or an atmosphere of composition like that of today, the amount of radiant energy received by Earth from the sun would not be enough to keep it from freezing. The way out of this dilemma is to have an atmosphere present during the early

Archean that was different in composition than that of today. . . . For a variety of reasons, it has been concluded, although still debated, that the most likely gases present in greater abundance in the Archean atmosphere were carbon dioxide, water vapor (the most important greenhouse gas) and perhaps methane. The presence of these greenhouse gases warmed the atmosphere and planetary surface and prevented the early Archean Earth from being frozen.[5]

(2) Find three inferences to the best explanation in the readings on scientific reasoning in Chapter 15. This should be easy because scientists use this form of argument often. Now, as in (1), put those inferences in standard form, and then evaluate them as carefully as you can, using the tests discussed above.

[5]From Fred T. Mackenzie, *Our Changing Planet* (Upper Saddle River, NJ: Prentice-Hall, 1998), 192.

312

ARGUMENTS FROM ANALOGY

Another common kind of inductive argument is an *argument from analogy.* Analogies and metaphors occur often in arguments, as we saw in Chapters 1 and 2. Arguments from analogy, however, use analogies in a specific way. They move from the premise that two things are similar in some respects to a conclusion that they must also be analogous in a further respect.

Such arguments can be found in many areas of everyday life. When we buy a new car, we can test drive it to see whether it is comfortable, and we can haggle about its price, but how can we tell whether it is going to be reliable? *Consumer Reports* might help if it is an old model; but if it is a brand-new model with no track record, then all we can go on is its similarities to earlier models. Our reasoning then seems to be that the new model is like the old model in various ways, and the old model was reliable, so the new model is probably reliable, too. Even *Consumer Reports* resorts to such arguments when they lack data on the reliability of a new model.

Although some people think that science should avoid analogies, the same form of argument is used in science. Here is an example from geology:

> Meteorites composed predominantly of iron provide evidence that parts of other bodies in the solar system, presumably similar in origin to Earth, were composed of metallic iron. The evidence from meteorite compositions and origins lends support to the conclusion that Earth's core is metallic iron.[6]

The argument here is that Earth is analogous to certain meteors in their origins, and those meteors have a large percentage of iron, so the Earth as a whole

[6]Mackenzie, *Our Changing Planet*, 42.

probably contains about the same percentage of iron. Because a smaller amount of iron is present in the Earth's crust, the rest must lie in the Earth's core.

In a different science, archaeologists might argue that a certain knife was used in ritual sacrifices because it resembles other sacrificial knives in its size, shape, materials, carvings, and so on. The analogy in this case is between the newly discovered knife and the other knives. This analogy is supposed to support a conclusion about the function of the newly discovered knife.

Although such arguments from analogy have diverse contents, they share a common form that can be represented like this:

(1) Object *A* has properties *P, Q, R*, and so on.
(2) Objects *B, C, D*, and so on also have properties *P, Q, R*, and so on.
(3) Objects *B, C, D*, and so on have property X.
∴ (4) Object *A* probably also has property X.

In the archaeological example, object *A* is the newly discovered knife, and objects *B, C, D*, and so on are previously discovered knives that are known to have been used in sacrifices. Properties *P, Q, R*, and so on are the size, shape, materials, and carvings that make *A* analogous to *B, C, D*, and so on. *X* is the property of being used as a sacrificial knife. Premise (3) says that the previously discovered artifacts have this property. The conclusion, (4), says that the newly discovered artifact probably also has this property.

Other arguments from analogy might refer to more or fewer objects, but at least two objects or kinds of objects are needed for an analogy. Arguments from analogy also might mention more or fewer properties or respects in which the objects are analogous. Some arguments might just say that the objects are analogous without specifying any particular respects at all.

Since arguments from analogy are inductive, they are usually not valid. It is possible that, even though this knife is analogous to other sacrificial knives, this knife was used to shave the king or just to cut bread. Such arguments are also defeasible. The argument about knives loses all of its strength if we find "Made in the U.S.A." printed on the newly discovered knife. But none of this shows that arguments from analogy are no good. Despite being inductive, some arguments from analogy can still provide reasons—even strong reasons—for their conclusions.

How can we tell whether an argument from analogy is strong or weak? One obvious requirement is that the premises must be *true.* If the previously discovered knives were not really used in sacrifices, or if they do not really have the same carvings on their handles as the newly discovered knife, then this argument from analogy does not provide much, if any, support for its conclusion.

Another requirement is that the cited similarities must be *relevant* and *important.* Suppose someone argues that his old car was red and had four doors and a sun roof, and his new car also has these properties, so his new car is probably going to be as reliable as his old car. This argument is very weak because the cited similarities are obviously irrelevant to reliability. Of course, features like this are almost never actually used in arguments from analogy. That is because people choose which properties to cite in their arguments from analogy in light of their background assumptions about what is relevant. Similar assumptions are needed when we assess someone else's argument from analogy. To determine which properties are relevant and important, we need to apply background beliefs, such as that reliability depends on the drive train and engine rather than on color or a sunroof. Without relying on some such background beliefs, it is impossible to construct or evaluate arguments from analogy.

If we are not sure which respects *are* important, we still might have some idea of which respects *might* be relevant; and then we can try to cite objects that are analogous in as many as possible of those respects. By increasing the number of potentially relevant respects for which the analogy holds, we can increase the likelihood that the important respects will be on our list. That shows why arguments from analogy are usually stronger when they cite *more and closer analogies* between the objects.

Another factor that affects the strength of an argument from analogy is the presence of *relevant disanalogies*. Because arguments from analogy are defeasible, as we saw, a strong argument from analogy can become weak if we add a premise that states an important disanalogy. Suppose my new car is like my old cars in many ways, but there is one difference: the new car has an electric motor, whereas the old cars were powered by gas. This one difference is enough to remove much (if not all) of the strength from any argument to the conclusion that the new car will be reliable. Of course, other disanalogies, such as a different color, won't matter to reliability; and it will often require background knowledge to determine how important a disanalogy is. What reduces the strength of an argument from analogy is important undermining disanalogies.

313

We need to be careful here. Some disanalogies that are relevant do not undermine an argument from analogy. If a new engine design was introduced by top engineers to increase reliability, then this disanalogy might not undermine the argument from analogy. Differences that point to more reliability rather than less might even make the argument from analogy stronger.

Other disanalogies can increase the strength of an argument from analogy in a different way. If the same markings are found on very different kinds of sacrificial knives, then the presence of those markings on the newly discovered knife is even stronger evidence that this knife was also used in sacrifices. Differences among the cases cited only in the premises as analogies (that is, *B*, *C*, *D*, and so on) can strengthen an argument from analogy.

Finally, the strength of an argument from analogy depends on its conclusion. Analogies to other kinds of cars provide stronger evidence for a weak conclusion (such as that the new model will probably be pretty reliable) and weaker evidence for a strong conclusion (such as that the new model will definitely be just as reliable as the old model). As with other forms of argument, an argument from analogy becomes stronger as its conclusion becomes weaker and vice versa.

These standards can be summarized by saying that an argument from analogy is stronger when:

(1) It cites more and closer analogies that are more important.
(2) There are fewer or less important disanalogies between the object in the conclusion and the other objects.
(3) The objects cited only in the premises are more diverse.
(4) The conclusion is weaker.

It will not always be easy to apply these criteria, because they often depend on background knowledge, but this list at least points us toward the main factors that need to be considered in evaluating an argument from analogy.

After learning about arguments from analogy, it is natural to wonder how they are related to inferences to the best explanation. Although this is sometimes disputed, it seems to us that arguments from analogy are often—if not always—implicit and incomplete inferences to the best explanation. As we pointed out, analogies don't support any conclusion unless they are relevant, and whether they are relevant depends on how they fit into explanations. The color of a car is irrelevant to its reliability, because color plays no role in explaining its reliability. What explains its reliability is its drive train design, materials, care in manufacturing, and so on. That is why analogies in those respects can support a conclusion about reliability. Similarly, the markings on an artifact are relevant to whether it is a sacrificial knife *if* the best explanation of why it has those markings is that it was used in sacrifices. What makes that explanation best is that it also explains similar markings on other sacrificial knives. Thus, such arguments from analogy can be seen as involving an inference to the best explanation of why objects *B*, *C*, *D*, and so on have property *X* followed by an application of that explanation to the newly discovered object *A*.

Sometimes the explanation runs in the other direction. Whereas the conclusion about the knife's use (*X*) is supposed to explain its shared markings (*P*, *Q*, *R*), sometimes it is the shared features (*P*, *Q*, *R*) that are supposed to explain the feature claimed in the conclusion (*X*). Here is a classic example:

> We may observe a very great [similarity] between this earth which we inhabit, and the other planets, Saturn, Jupiter, Mars, Venus, and Mercury. They all revolve around the sun, as the earth does, although at different distances and in different periods. They borrow all their light from the sun, as the earth does. Several of them are known to revolve around their axis like the earth, and, by that means, must have a like succession of day and night. Some of them have moons that serve to give them light in the absence of the sun, as our moon does to us. They are all, in their motions, subject to the same law of gravitation, as the earth is. From all this similarity it is not unreasonable to think that those planets may, like our earth, be the habitation of various orders of living creatures. There is some probability in this conclusion from analogy.[7]

[7]Thomas Reid, *Essays on the Intellectual Powers of Man* (Cambridge, MA: MIT Press, 1969), essay I, section 4, 48.

The argument here seems to be that some other planet probably supports life, because Earth does and other planets are similar to Earth in revolving around the sun and around an axis, getting light from the sun, and so on. Different analogies would be added in more recent versions of this argument, but its form remains the same. What makes certain analogies relevant is not, of course, that the motion of Earth is explained by the presence of life here. Rather, certain features of Earth explain why Earth is habitable. The argument suggests that the best explanation of why there is life on our planet is that certain conditions make life possible. That generalization can then be used to support the conclusion that other planets with the same conditions probably support life as well.

Thus, in one way or another, arguments from analogy can be seen as inferences to the best explanation. But they are usually incomplete explanations. The argument for life on other planets did not have to commit itself to any particular theory about the origin of life or about which conditions are needed to support life. Nor did the car argument specify exactly what makes cars reliable. Such arguments from analogy merely list a number of similarities so that the list will be likely to include whatever factors are needed for life or for reliability. In this way, arguments from analogy can avoid depending on any complete theory about what is and what is not relevant.

This incompleteness makes arguments from analogy useful in situations where we do not yet know enough to formulate detailed theories or even to complete an inference to the best explanation. However, the incompleteness of arguments from analogy also makes them more vulnerable to refutation, since the analogies that they list might fail to include a crucial respect. As with other kinds of arguments, the very features that make a form of argument available in a wider variety of situations also makes it less secure. This does not mean that arguments from analogy are never any good. They can be strong. But it does suggest that their strength will increase as they approach or approximate more complete inferences to the best explanation.[8]

EXERCISE IV

For each of the following arguments, state whether the indicated changes would make the argument weaker or stronger, and explain why. The strength of the argument might not be affected at all. If so, say why it is not affected.

(1) My friend and I have seen many movies together, and we have always agreed on whether they are good or bad. My friend liked the movie *Pulp Fiction*. So I probably will like *Pulp Fiction* as well.

[8] For more on arguments from analogy, see "Precedents" in Chapter 13 and "Analogical Reasoning in Ethics" in Chapter 14.

Would this argument be weaker or stronger if:

(a) The only movies that my friend and I have watched together are comedies, and *Pulp Fiction* is not a comedy.
(b) My friend and I have seen very many, very different movies together.
(c) My friend and I always watched movies together on Fridays, but my friend watched *Pulp Fiction* on a Saturday.
(d) The conclusion claims that I definitely will like *Pulp Fiction*.
(e) The conclusion claims that I probably won't think that *Pulp Fiction* is total trash.

(2) All the students from this high school with high grades and high board scores did well in college. Joe also had high grades and board scores. So he will probably do well in college.

Would this argument be weaker or stronger if:

(a) Joe is lazy, but the other students worked hard.
(b) Joe got good grades and board scores, but he did not go to the same high school as the students with whom he is being compared.
(c) Joe also got good grades, but he did not get good board scores or go to the same high school as the students with whom he is being compared.
(d) Joe is going to a different college than the students with whom he is being compared.
(e) Joe is premed, but the other students are majoring in physical education.
(f) Joe is majoring in physical education, but the other students are premed.
(g) The conclusion is that Joe will graduate first in his class.

(3) A new drug cures a serious disease in rats. Rats are similar to humans in many respects. Therefore, the drug will probably cure the same disease in humans.

Would this argument be weaker or stronger if:

(a) The disease affects the liver, and rat livers are very similar to human livers.
(b) The drug does not cure this disease in cats.
(c) The drug has to be injected into the rat's tail to be effective.
(d) No drug of this general type has been used on humans before.

EXERCISE V

Using the criteria mentioned above, evaluate each of the following arguments as strong or weak. Explain your answers. Be sure to specify the properties on which the analogy is based, as well as any background beliefs on which your evaluation depends.

(1) This landscape by Cézanne is beautiful. He did another painting of a similar scene around the same time. So it is probably beautiful, too.
(2) My aunt had a Siamese cat that bit me, so this Siamese cat will probably bite me, too.

(3) The students I know who took this course last year got grades of A. I am a lot like them, since I am also smart and hardworking; and the course this year covers very similar material. So I will probably get an A.

(4) This politician was convicted of cheating in his marriage, and he will have to face similarly strong temptations in his public duties, so he will probably cheat in political life as well.

(5) A very high minimum wage led to increased unemployment in one country. That country's economy is similar to the economy in a different country. So a very high minimum wage will probably lead to increased unemployment in the other country as well.

(6) I feel pain when someone hits me hard on the head with a baseball bat. Your body is a lot like mine. So you would probably feel pain if I hit you hard on the head with a baseball bat. (This is related to the "Problem of Other Minds.")

(7) It is immoral for a doctor to lie to a patient about a test result, even if the doctor thinks that lying is in the patient's best interest. We know this because even doctors would agree that it would be morally wrong for a financial advisor to lie to them about a potential investment, even if the financial advisor thinks that this lie is in the doctor's best interests.

(8) Chrysler was held legally liable for damages due to defects in the suspension of its Corvair. The defects in the Pinto gas tank caused injuries that were just as serious. Thus, Ford should also be held legally liable for damages due to those defects.

315

EXERCISE VI

More detailed examples of arguments from analogy can be found in Part 2.

(1) In Chapter 13, critics argue that preferential treatment is unconstitutional because of analogies to other forms of racial discrimination that were found unconstitutional in precedents.

(2) In Chapter 14, Thomson defends the morality of abortion by means of an analogy to a kidnapped violinist.

(3) In Chapter 15, Alvarez and Asaro argue that some material in the KT boundary clay probably came from outer space because it is similar in composition to materials found in meteors.

(4) In Chapter 16, Searle argues that computers cannot think just by virtue of their formal structure, because of an analogy to a Chinese room with the same formal structure.

Evaluate each of these arguments by applying the criteria discussed above.

DISCUSSION QUESTIONS

(1) Paley's argument from design for the existence of God, given in the following excerpt, is often presented as a classic example of an argument from analogy. Is it really an argument from analogy, an inference to the best explanation, or both? Why? You might try to reconstruct the argument and put it into standard form before you try to classify it.

> In crossing a heath, suppose I pitched my foot against a *stone*, and were asked how the stone came to be there; I might possibly answer, that, for anything I knew to the contrary, it had lain there forever: nor would it, perhaps, be very easy to show the absurdity of this answer. But suppose I had found a *watch* upon the ground, and it should be inquired how the watch happened to be in that place. . . . When we come to inspect the watch, we perceive—what we could not discover in the stone—that its several parts are framed and put together for a purpose. . . . This mechanism being observed . . . the inference we think is inevitable, that the watch must have had a maker. . . . Every indication of contrivance, every manifestation of design, which existed in the watch, exists in the works of nature, with the difference on the side of nature of being greater and more. . . . Were there no example in the world of contrivance except that of the *eye*, it would be alone sufficient to support the conclusion which we draw from it, as to the necessity of an intelligent Creator. . . . Its coats and humors, constructed as the lenses of a telescope are constructed, for the refraction of rays of light to a point, which forms the proper action of the organ; the provision in its muscular tendons for turning its pupil to the object, similar to that which is given to the telescope by screws, and upon which power of direction in the eye the exercise of its office as an optical instrument depends; . . . these provisions compose altogether an apparatus, a system of parts, a preparation of means, so manifest in their design, so exquisite in their contrivance, so successful in their issue, so precious, and so infinitely beneficial in their use, as, in my opinion, to bear down all doubt that can be raised on the subject.[9]

(2) The following excerpt comes from a newspaper report in the *Toronto Star*, October 10, 1999. Put the central argument from analogy, which is italicized, into standard form. Then reconstruct the argument as an inference to the best explanation. Which representation best captures the force of the argument, or are they equally good?

A GNAWING QUESTION IS ANSWERED

by Michael Downey

Tim White is worried that he may have helped to pin a bad rap on the Neanderthals, the prehistoric Europeans who died out 25,000 years ago. "There is a danger that everyone will think that all Neanderthals were cannibals and that's not necessarily true," he says. White was part of a French-American team of paleoanthropologists who recently found conclusive evidence that at

[9] William Paley, *Natural Theology* (New York: Bobbs-Merrill, 1963), 3, 4, 13, 32.

least some Neanderthals ate others about 100,000 years ago. But that doesn't mean they were cannibalistic by nature, he stresses. Most people don't realize that cannibalism is widespread throughout nature, says White, a professor at the University of California at Berkeley and the author of a book on prehistoric cannibalism.

The question of whether the Neanderthals were cannibals had long been a hotly debated topic among anthropologists. No proof had ever been found. That debate ended, however, with the recent analysis by the team of stone tools and bones found in a cave at Moula-Guercy in southern France. The cave is about the size of a living room, perched about 80 metres above the Rhone River. "This one site has all of the evidence right together," says White. "It's as if somebody put a yellow tape around the cave for 100,000 years and kept the scene intact." The bones of deer and other fauna show the clear markings of the nearby stone tools, indicating the deer had been expertly butchered; they were skinned, their body parts cut off and the meat and tendons sliced from the bone. Long bones were bashed open "to get at the fatty marrow inside," says White.

So what does all this have to do with cannibalism? *The bones of the six (so far) humans in the same location have precisely the same markings made by the same tools. That means these fairly modern humans were skinned and eaten in the same manner as the deer.*

And if you are thinking they were eaten after they just happened to die, they do represent all age groups. Two were children about 6 years old, two were teenagers and two were adults.

But maybe they were eaten at a time when food was unusually scarce, right? Not so. There is a large number of animal bones at the same dig, indicating that there were options to eating other Neanderthals.

Human bones with similar cut marks have been found throughout Europe, from Spain to Croatia, providing tantalizing hints of Neanderthal cannibalism activity over tens of thousands of years. But finding such clear evidence of the same preparation techniques being used on deer in the same cave site in France, will "necessitate reassessment of earlier finds," always attributed to ritual burial practices or some other explanation, says White.

It was not clear whether the Neanderthals ate human flesh of their own tribe or exclusively from an enemy tribe, White stresses. Nor was there any indication the purpose of the cannibalism involved nourishment. Eating human flesh could have had another purpose altogether, he says. Surprising to some, cannibalism has been found in 75 mammal species and in 15 primate groups. White says it has often been practised for reasons not associated with normal hunger. White quotes an archeological maxim: "Actions fossilize, intentions don't." In other words, the reason for the cannibalism remains unknown. He notes that the flesh of other humans has sometimes been eaten to stave off starvation, to show contempt for an enemy or as part of a ritual of affection for the deceased. "Were the victims already dead or killed to be eaten?" he asks, "Were they enemies of the tribe or family members?" Learning these details is the "important and extremely difficult part."

This excavation represents a breakthrough in archeological practice. In a series of papers, White has long advocated the importance of treating a dig site like a crime scene, leaving every piece of evidence in place. In many earlier digs, animal bones were frequently pulled out and thrown away as being irrelevant and human bones were often coated with shellac. The human bones were all tossed into the same bag with no regard to their juxtaposition to each other or precise location. This is one of the first times that modern forensic techniques have been utilized in an archeological excavation, White says, and conclusions drawn have been much more precise than in previous digs that used cruder methods. The project team, which is headed by Alban Defleur of the Universite du Mediterrane at Marseilles, has been digging in the cave since 1991.

316

REASONING ABOUT CAUSES

Many explanations depend on causal generalizations. If our car goes dead in the middle of rush-hour traffic just after its 20,000-mile checkup, we assume that there must be some reason why this happened. Cars just don't stop for no reason at all. So we ask: "What caused our car to stop?" The answer might be that it ran out of gas. If we find, in fact, that it did run out of gas, then that will usually be the end of the matter. We will think that we have discovered why this particular car stopped running. This reasoning is about a particular car on a particular occasion, but it rests on certain *generalizations:* we are confident that *our* car stopped running when it ran out of gas, because we believe that *all* cars stop running when they run out of gas. We probably did not think about this, but our causal reasoning in this particular case appealed to a commonly accepted *causal generalization:* cars will not run without gas.

Causal generalizations are also used to *predict* the consequences of particular actions or events. A race car driver might wonder, for example, what would happen if he added just a bit of nitroglycerin to his fuel mixture. Would it give him better acceleration, blow him up, do very little, or what? In fact, the driver may not be in a position to answer this question straight off, but his thinking will be guided by the causal generalization that igniting nitroglycerin can cause a dangerous explosion.

So a similar pattern arises for both causal explanation and causal prediction. These inferences contain two essential elements:

(1) The facts in the particular case. (For example, the car stopped and the gas gauge reads empty, or I just put a pint of nitroglycerine in the gas tank of my Maserati, and I am about to turn the ignition key.)

(2) Certain causal generalizations. (For example, cars do not run without gas or nitroglycerin explodes when ignited.)

The basic idea is that causal inferences bring particular facts under causal generalizations.

This shows why causal generalizations are important, but what exactly are they? Although this issue remains controversial, here we will treat them as a kind of *general conditional*. A general conditional has the following form:

For all x, if x has the feature F, then x has the feature G.

We will say that, according to this conditional, x's having the feature F is a *sufficient condition* for its having the feature G; and x's having the feature G is a *necessary condition* for its having the feature F.

Some general conditionals are not *causal*. Neither of these two general conditionals expresses a causal relationship:

If something is a square, then it is a rectangle.

If you are eighteen years old, then you are eligible to vote.

The first conditional tells us that being a square is sufficient for being a rectangle, but this is a mathematical (or a priori) relationship, not a causal one. The second conditional tells us that being eighteen years old is a sufficient condition for being eligible to vote. The relationship here is legal, not causal.

Although many general conditionals are not causal, all causal conditionals are general, in our view. Consequently, if we are able to show that a causal conditional is false just by virtue of its being a general conditional, we will have refuted it. This will serve our purposes well, for in what follows we will be largely concerned with finding reasons for *rejecting* causal generalizations.

It is important to weed out false causal generalizations, because they can create lots of trouble. Doctors used to think that bloodletting would cure disease. They killed many people in the process of trying to cure them. Thus, although we need causal generalizations for getting along in the world, we also need to get them right. We will be more likely to succeed if we have proper principles for testing and applying such generalizations. It is one of the central tasks of inductive logic to supply these principles.

In the past, very elaborate procedures have been developed for this purpose. The most famous set of such procedures was developed by John Stuart Mill and has come to be known as Mill's methods.[10] Though inspired by Mill's methods, the procedures introduced here involve some fundamental simplifications; whereas Mill introduced five methods, we will introduce only three primary rules.

The first two rules are the Sufficient Condition Test (SCT) and the Necessary Condition Test (NCT). We will introduce these tests first at an abstract level. One advantage of formulating these tests abstractly is so that they can be applied to other kinds of sufficient and necessary conditions, for example, those that arise in legal and moral reasoning, the topics of Chapters 13 and

[10] Mill's "methods of experimental enquiry" are found in book 3, chap. 8 of his *A System of Logic* (London: John W. Parker, 1843). Our simplification of Mill's methods derives from Brian Skyrms, *Choice and Chance*, 3d ed. (Belmont, CA: Wadsworth, 1986), chap. 4.

14. Once it is clear how these tests work in general, we will apply them specifically to causal reasoning.

Sufficient Conditions and Necessary Conditions

To keep our discussion as general as possible, we will adopt the following definitions of sufficient conditions and necessary conditions:

Feature F is a *sufficient* condition for feature G if and only if anything that has feature F also has feature G.

Feature F is a *necessary* condition for feature G if and only if anything that *lacks* feature F also *lacks* feature G.

These definitions are equivalent to those in the previous section, because, if anything that *lacks* feature F also *lacks* feature G, then anything that *has* feature G must also *have* feature F; and if anything that *has* feature G must also *have* feature F, then anything that *lacks* feature F also *lacks* feature G. It follows that feature F is a sufficient condition for feature G if and only if feature G is a necessary condition for feature F.[11]

It is important not to confuse sufficient conditions with necessary conditions. Something can be a sufficient condition for a feature without being a necessary condition for that feature. For example, being a mother is a sufficient condition for being female, but it is not a necessary condition for being female. Furthermore, something can be a necessary condition for a feature without being a sufficient condition for that feature. Although being a mother is a necessary condition for being a grandmother, it is not a sufficient condition for being a grandmother. Of course, some necessary conditions are also sufficient conditions. Being a mother is both necessary and sufficient for being a female parent. Nonetheless, because many necessary conditions are not sufficient conditions, and many sufficient conditions are not necessary conditions, we need to distinguish the two kinds of conditions.

This distinction becomes complicated when conditions get complex. Our definitions and tests hold for all features, whether positive or negative (such as having hair or not having hair) and whether simple or conjunctive (such as having both a beard and a mustache) or disjunctive (such as having either a beard or a mustache). Thus, not having any hair (anywhere) on your head is a sufficient condition of not having a beard, so not having a beard is a necessary condition of not having any hair on your head. However, not having any hair on your head is not necessary for not having a beard, since you can have some hair on the top of your head without having a beard. Negation can create confusion, so we need to think carefully about what is being claimed to be necessary or sufficient for what.

[11] Symbolization might make this clearer, because in quantificational logic "$(x)(Fx \supset Gx)$" is equivalent to "$(x)(\sim Gx \supset \sim Fx)$" (see Chapter 8).

Even in simple cases, there is a widespread tendency to confuse necessary conditions with sufficient conditions. One reason is that, as we saw in Chapter 6, conditionals often conversationally imply biconditionals. For example, suppose a student wants to take an advanced philosophy seminar next year, so she asks the professor what the prerequisites are. The professor answers, "Students can take my seminar next year only if they have taken a course in ethics before the seminar starts." If the student takes a course in ethics but finds out too late that other courses are also required, then the student would rightfully feel cheated. Why? What the professor said was literally true: a course in moral philosophy is a necessary condition of eligibility for the seminar. Nonetheless, it was misleading for the professor to say this alone, because he violated Grice's conversational rule of Quantity. As in other cooperative exchanges, in stating necessary conditions we are expected to specify *all* of the necessary conditions that we know to be relevant in the given context (except, perhaps, those that are so obvious that they can be taken for granted). This explains why a statement that certain conditions are necessary often conversationally implies that these are all of the necessary conditions. It is often plausible to assume that all necessary conditions taken together will be a sufficient condition. These two steps together explain why a statement of necessary conditions will often be taken as a statement of a sufficient condition.

318

Seeing how it is natural for us to confuse sufficient conditions and necessary conditions can help us guard against doing so. It is important to keep these concepts straight, for, as we will see, the tests concerning them are fundamentally different.

EXERCISE VII

Which of the following claims are true? Which are false?

(1) Being a car is a sufficient condition for being a vehicle.
(2) Being a car is a necessary condition for being a vehicle.
(3) Being a vehicle is a sufficient condition for being a car.
(4) Being a vehicle is a necessary condition for being a car.
(5) Being an integer is a sufficient condition for being an even number.
(6) Being an integer is a necessary condition for being an even number.
(7) Being an integer is a sufficient condition for being either even or odd.
(8) Being an integer is a necessary condition for being either even or odd.
(9) Not being an integer is a sufficient condition for not being odd.
(10) Not being an integer is a sufficient condition for not being even.
(11) Being both an integer and divisible by two without remainder is a sufficient condition for being even.
(12) Being both an integer and divisible by two without remainder is a necessary condition for being even.

(13) Being divisible by two without remainder is a necessary condition for being even.
(14) Driving seventy-five miles per hour (for fun) is a sufficient condition for violating a legal speed limit of sixty-five miles per hour.
(15) Driving seventy-five miles per hour (for fun) is a necessary condition for violating a legal speed limit of sixty-five miles per hour.
(16) Cutting off Joe's head is a sufficient condition for killing him.
(17) Cutting off Joe's head is a necessary condition for killing him.
(18) Cutting off Joe's head and then holding his head under water for ten minutes is a sufficient condition for killing him.

EXERCISE VIII

Using the above definitions, indicate whether the following sentences claim that the underlined words express a necessary condition or a sufficient condition of the italicized words.

EXAMPLE: If <u>you work hard,</u> *you will succeed.*
SOLUTION: Your working hard is sufficient for your success.

(1) *Litmus paper will turn red* if <u>it is put in acid.</u>
(2) *Litmus paper will turn red* only if <u>it is put in acid.</u>
(3) *Litmus paper will not turn red* if <u>it is not put in acid.</u>
(4) <u>Litmus paper will turn red</u> if *it is put in acid.*
(5) <u>Litmus paper will not turn red</u> provided that *it is not put in acid.*
(6) <u>You have to pass the final exam</u> in order to *pass the course.*
(7) <u>Passing the final exam</u> is all you have to do to *pass the course.*
(8) If you don't <u>pass the final exam,</u> you can't *pass the course.*
(9) If you <u>pass the final exam,</u> you can't fail to *pass the course.*
(10) You can't *pass this course* unless you <u>pass the final exam.</u>

EXERCISE IX

Indicate whether the following principles are true or false and why.

(1) If having feature *F* is a *sufficient* condition for having feature *G*, then having feature *G* is a *necessary* condition for having feature *F*.
(2) If having feature *F* is a *sufficient* condition for having feature *G*, then lacking feature *F* is a *necessary* condition for lacking feature *G*.
(3) If lacking feature *F* is a *sufficient* condition for having feature *G*, then having feature *F* is a *necessary* condition for lacking feature *G*.

(4) If having feature *F* is a *sufficient* condition for having feature *G*, then lacking feature *F* is a *necessary* condition for lacking feature *G*.
(5) If having either feature *F* or feature *G* is a *sufficient* condition for having feature *H*, then having feature *F* is a *sufficient* condition for having feature *H*.
(6) If having either feature *F* or feature *G* is a *sufficient* condition for having feature *H*, then having feature *G* is a *sufficient* condition for having feature *H*.
(7) If having either feature *F* or feature *G* is a *sufficient* condition for having feature *H*, then not having feature *F* is a *necessary* condition for not having feature *H*.
(8) If having both feature *F* and feature *G* is a *necessary* condition for having feature *H*, then lacking feature *F* is a *sufficient* condition for lacking feature *H*.
(9) If not having both feature *F* and feature *G* is a *sufficient* condition for having feature *H*, then lacking feature *F* is a *sufficient* condition for having feature *H*.
(10) If having either feature *F* or feature *G* is a *sufficient* condition for having feature *H*, then having both feature *F* and feature *G* is a *sufficient* condition for having feature *H*.

19

THE SUFFICIENT CONDITION TEST[12]

We can now formulate tests to determine when something meets our definitions of sufficient conditions and necessary conditions. It will simplify matters if we first state these tests formally using letters. We will also begin with a simple case where we consider only four *candidates*—*A*, *B*, *C*, and *D*—for sufficient conditions for a *target* feature, *G*. *A* will indicate that the feature is present; ~*A* will indicate that this feature is absent. Using these conventions, suppose that we are trying to decide whether any of the four features—*A*, *B*, *C*, or *D*—could be a sufficient condition for *G*. To this end we collect data of the following kind:

TABLE 1

Case 1:	*A*	*B*	*C*	*D*	*G*
Case 2:	~*A*	*B*	*C*	~*D*	~*G*
Case 3:	*A*	~*B*	~*C*	~*D*	~*G*

We know by definition that, for something to be a sufficient condition of something else, when the former is present, the latter must be present as well. Thus, to test whether a candidate really is a sufficient condition of *G*, we only have to examine cases where the target feature, *G*, is absent, and then check to see whether any of the candidate features are present.

> SCT: Any candidate that is present when *G* is absent is eliminated as a possible sufficient condition.

The test applies to Table 1 as follows: Case 1 need not be examined because *G* is present, so there can be no violation of SCT in Case 1. Case 2 eliminates two of the candidates, *B* and *C*, for both are present in a situation in which *G* is absent. Finally, Case 3 eliminates *A* for the same reason. We are thus left with *D* as our only remaining candidate for a sufficient condition for *G*.

Now let's consider *D*. Having survived the application of the SCT, does it follow that *D* is a sufficient condition for *G*? No! On the basis of what we have been told so far, it remains entirely possible that the discovery of a further case will reveal an instance where *D* is present and *G* absent, thus showing that *D* is not a sufficient condition for *G* either.

Case 4:	*A*	*B*	~*C*	*D*	~*G*

In this way, it is always possible for new cases to refute any inference from a limited group of cases to the conclusion that a certain candidate is a sufficient condition. In contrast, no further case can change the fact that *A*, *B*, and *C* are not sufficient conditions, because they fail the SCT.

This shows that, when we apply the SCT to rule out a candidate as a sufficient condition, our argument is *deductive*. We simply find a counterexample to the universal claim that a certain feature is sufficient. However, when a candidate is not ruled out and we draw the positive conclusion that that candidate *is* a sufficient condition, then our argument is *inductive*. Inductive inferences, however well confirmed, are always defeasible. (Remember Captain Cook's discovery of black swans.) That is why our inductive inference to the conclusion that *D* is a sufficient condition could be refuted by the new case, 4.

THE NECESSARY CONDITION TEST[13]

The necessary condition test (NCT) is like the SCT but works in the reverse fashion. With SCT we eliminated a candidate *F* from being the sufficient condition for *G*, if *F* was ever present when *G* was absent. With the necessary condition test, we eliminate a candidate *F* from being a necessary condition for *G* if we can find a case where *G* is present, but *F* is not. This makes sense, because if *G* can be present when *F* is not, then *F* cannot be necessary for the occurrence of *G*. Thus, in applying the necessary condition test, we only have to examine cases where the target feature, *G*, is present, and then check to see whether any of the candidate features are absent.

> NCT: Any candidate that is absent when *G* is present is eliminated as a possible necessary condition of *G*.

The following table gives an example of an application of this test:

TABLE 2:

Case 1:	*A*	*B*	*C*	*D*	~*G*
Case 2:	~*A*	*B*	*C*	*D*	*G*
Case 3:	*A*	~*B*	*C*	~*D*	*G*

[12]This test parallels but is not identical with Mill's Method of Difference.

[13]This test parallels but is not identical with Mill's Method of Agreement.

Because Case 1 does not provide an instance where G is present, it cannot eliminate any candidate as a necessary condition. Case 2 eliminates A, since it shows that G can be present without A being present. Case 3 eliminates both B and D, leaving C as the only possible candidate for being a necessary condition for G. From this, of course, it does not follow that C *is* a necessary condition for G, for, as always, new cases might eliminate it as well. The situation is the same as with the SCT. An argument for a negative conclusion that a candidate is not a necessary condition, because that candidate fails the NCT, is a deductive argument that cannot be overturned by any further cases. In contrast, an argument for a positive conclusion that a candidate is a necessary condition because that candidate fails the NCT is an inductive argument that can be overturned by a further case where this candidate fails the NCT.

THE JOINT TEST[14]

It is also possible to apply these rules simultaneously in the search for possible conditions that are both sufficient and necessary. Any candidate cannot be both sufficient and necessary if it fails either the SCT or the NCT. In Table 2, C is the only possible necessary condition for G, and it is not also a possible sufficient condition, since C fails the SCT in Case 1, where C is present and G is absent. In Table 1, however, D is a possible sufficient condition of G, because D is never present when G is absent; and D might also be a necessary condition for G, since G is never present when D is absent. Thus, none of the cases in Table 1 eliminates D as a candidate for a condition that is both sufficient and necessary. As before, this possibility still might be refuted by Case 4, so any inference to a positive conclusion that some candidate is a necessary and sufficient condition must be defeasible and, hence, inductive.

EXERCISE X

For each of the following tables decide:

(a) Which, if any, of the candidates—A, B, C, or D—is not eliminated by the sufficient condition test.
(b) Which, if any, of the candidates—A, B, C, or D—is not eliminated by the necessary condition test.
(c) Which, if any, of the candidates—A, B, C, or D—is not eliminated by either test.

EXAMPLE:	Case 1.	A	B	$\sim C$	D	$\sim G$
	Case 2.	$\sim A$	B	C	D	G
	Case 3.	A	$\sim B$	C	D	G

[14]This test parallels but is not identical with Mill's Joint Method of Agreement and Difference.

(a) Only C passes the SCT.
(b) C and D both pass the NCT.
(c) Only C passes both.

(1)	Case 1:	A	B	C	D	G
	Case 2:	$\sim A$	B	$\sim C$	D	$\sim G$
	Case 3:	A	$\sim B$	C	$\sim D$	G
(2)	Case 1:	A	B	C	$\sim D$	G
	Case 2:	$\sim A$	B	C	D	G
	Case 3:	A	$\sim B$	C	$\sim D$	G
(3)	Case 1:	A	B	C	D	$\sim G$
	Case 2:	$\sim A$	B	C	D	G
	Case 3:	A	$\sim B$	C	$\sim D$	G

EXERCISE XI

Suppose your computer won't work, and you want to find out why. After checking to make sure that it is plugged in, you experiment with a new computer, a new (external) hard disk, and new system software in the combinations on the table below.

	Plug	Computer	Hard Disk	Software	Result
Case 1	In	Old Mac	Old HD	Old SW	Fails
Case 2	In	Old Mac	Old HD	New SW	Works
Case 3	In	Old Mac	New HD	Old SW	Fails
Case 4	In	Old Mac	New HD	New SW	Works
Case 5	In	New Mac	Old HD	Old SW	Works
Case 6	In	New Mac	Old HD	New SW	Works
Case 7	In	New Mac	New HD	Old SW	Fails
Case 8	In	New Mac	New HD	New SW	Works

The candidates for necessary conditions and sufficient conditions are the plug position, the computer, the hard disk, and the software. For each candidate, say (a) which cases, if any, eliminate it as a sufficient condition of your computer's failure and (b) which cases, if any, eliminate it as a necessary condition of your computer's failure. Which candidates, if any, are not eliminated as a sufficient condition? As a necessary condition? Does it follow that these candidates are necessary conditions or sufficient conditions? Why or why not?

RIGOROUS TESTING

Going back to Table 1, it is easy to see that A, B, C, and D are not eliminated by NCT as necessary conditions of G, as G is only present in one case (Case 1) and A, B, C, and D are present there as well. So far so good; but if we wanted to test these features more rigorously, it would be important to find more cases in which *G* was present and see if these candidates are also present and thus continue to survive the NCT.

The following table gives a more extreme example of nonrigorous testing:

TABLE 3

Case 1:	*A*	~*B*	*C*	*D*	*G*
Case 2:	*A*	~*B*	~*C*	~*D*	~*G*
Case 3:	*A*	~*B*	*C*	~*D*	~*G*
Case 4:	*A*	~*B*	~*C*	*D*	*G*

Here, *A* is eliminated by SCT (in Cases 2 and 3) but is not eliminated by NCT, so it is a possible necessary condition but not a possible sufficient condition. *B* is not eliminated by SCT but is eliminated by NCT (in Cases 1 and 4), so it is a possible sufficient condition but not a possible necessary condition. *C* is eliminated by both rules (in Cases 3 and 4). Only *D* is not eliminated by either test, so it is the only candidate for being both a necessary and a sufficient condition for *G*.

321

The peculiarity of this example is that *A* is always present whether *G* is present or not, and *B* is always absent whether *G* is absent or not. Now if something is always present, as *A* is, then it cannot possibly fail the NCT; for there cannot be a case where the target is present and the candidate is absent if the candidate is *always* present. If we want to test *A* rigorously under the NCT, then we should try to find cases in which *A* is absent and then check to see whether *G* is absent as well.

In reverse fashion, but for similar reasons, if we want to test *B* rigorously under the SCT, we should try to find cases in which *B* is present and then check to see if *G* is present as well. If we restrict our attention to cases where *B* is always absent, as in Table 3, then *B* cannot fail the SCT, but passing that test will be trivial for *B* and so will not even begin to show that *B* is a sufficient condition.

Now consider two more sets of data just like Table 2, except with regard to the target feature, *G*:

TABLE 4:

Case 1:	*A*	*B*	*C*	*D*	*G*
Case 2:	~*A*	*B*	*C*	*D*	*G*
Case 3:	*A*	~*B*	*C*	~*D*	*G*

TABLE 5:

Case 1:	*A*	*B*	*C*	*D*	~*G*
Case 2:	~*A*	*B*	*C*	*D*	~*G*
Case 3:	*A*	~*B*	*C*	~*D*	~*G*

Because *G* is present in all of the cases in Table 4, no candidates can be eliminated by the SCT. This result is trivial, however. Table 4 does not provide rigorous testing for a sufficient condition, because our attention is restricted to

a range of cases that is too narrow. Nothing could possibly be eliminated as a sufficient condition as long as *G* is always present. . . .

Similarly, *G* is absent in all of the cases in Table 5, so no candidates can be eliminated by the NCT. Still, because this data is so limited, its failure to eliminate candidates does not even begin to show that anything is a necessary condition.

For both rules, then, rigorous testing involves seeking out cases where failing the test is a live possibility. For the SCT, this requires looking both at cases in which the candidates are present and also at cases where the target is absent. For the NCT, rigorous testing requires looking both at cases in which the candidates are absent and also at cases where the target is present. Without cases like these, passing the tests is rather like a person bragging that he has never struck out when, in fact, he has never come up to bat.

Reaching Positive Conclusions

Suppose that we performed rigorous testing on candidate *C*, and it passed the SCT with flying colors. Can we now draw the positive conclusion that *C* is a sufficient condition? That depends on which kinds of candidates and cases have been considered. Since rigorous testing was passed, these three conditions are met:

(a) We have not found any case where the candidate, *C*, is present and the target, *G*, is absent.

(b) We have tested some cases where the candidate, *C*, is present.

(c) We have tested some cases where the target, *G*, is absent.

For it to be reasonable to reach a positive conclusion that *C* is sufficient for *G*, this further condition must also be met:

(d) We have tested enough cases of the various kinds that are likely to include a case where *C* is present and *G* is absent if there is any such case.

This new condition cannot be applied in the mechanical way that conditions (a)–(c) could be applied. To determine whether condition (d) is met, we need to rely on *background information* about how many cases are "enough" and about which "kinds of cases are likely to include a case where *C* is present and *G* is absent if there is any such case." For example, if we are trying to figure out whether our software is causing our computer to crash, we do not need to try the same kind of computer in different colors. We may assume that strawberry-colored iMacs are no more or less likely to crash than blueberry-colored ones. What we need to try are different kinds of computers, hard drives, monitors, plug positions, and so on, because we know that these are the kinds of factors that can affect performance. Background information like this is what tells us when we have tested enough cases of the right kinds.

Of course, our background assumptions might turn out to be wrong. Even if we have tested many variations of every factor that we think might be relevant, we still might be surprised and find a further case where C and ~G are present. However, all that shows is that our inference is defeasible, like all inductive arguments. Despite the possibility that future discoveries might undermine it, our inductive inference can still be strong if our background beliefs are justified and if we have looked long and hard without finding any case where C is present and G is absent.

Similar rules apply in reverse to positive conclusions about necessary conditions. We have good reason to suppose that candidate C is a necessary condition for target G, if the following conditions are met:

(a) We have not found any case where the candidate, C, is absent and the target, G, is present.

(b) We have tested some cases where the candidate, C, is absent.

(c) We have tested some cases where the target, G, is present.

(d) We have tested enough cases of the various kinds that are likely to include a case where C is absent and G is present if there is any such case.

As before, this argument is defeasible, but, if our background assumptions are justified, the fact that conditions (a)–(d) are all met can still provide a strong reason for the positive conclusion that C is a necessary condition for G.

222

The SCT and NCT themselves are still negative and deductive. However, that does not make them better than the positive tests encapsulated in conditions (a)–(d). The negative SCT and NCT are of no use when we need to argue that some condition is sufficient or is necessary. Such positive conclusions can be reached only by applying something like condition (d), which will require relying on background information. These inductive arguments might not be as clear-cut or secure as the negative ones, but they can still be inductively strong under the right circumstances; and that is all they claim to be.

Applying These Methods to Find Causes

In stating the SCT and NCT and applying these tests to abstract patterns of conditions to eliminate candidates, our procedure was fairly mechanical. We cannot be so mechanical when we try to reach positive conclusions that certain conditions are necessary, sufficient, or both. Applying these rules to actual concrete situations introduces a number of further complications, especially when using our tests to determine causes.

Normality. First, it is important to keep in mind that in our ordinary understanding of causal conditions, we usually take it for granted that the setting is normal. It is part of common knowledge that if you strike a match, then it will light. Thus, we consider striking a match sufficient to make it light. But if someone has filled the room with carbon dioxide, then the match will not light, no matter how it is struck. Here one may be inclined to say that, after all, striking a match is not sufficient to light it. We might try to be more careful and say that if a match is struck *and* the room is not filled with

carbon dioxide, then it will light. But this new conditional overlooks other possibilities—for example, that the room has been filled with nitrogen, that the match has been fireproofed, that the wrong end of the match was struck, that the match has already been lit, and so forth. It now seems that the antecedent of our conditional will have to be endlessly long in order to specify a true or genuine sufficient condition. In fact, however, we usually feel quite happy with saying that if you strike a match, then it will light. We simply do not worry about the possibility that the room has been filled with carbon dioxide, the match has been fireproofed, and so on. Normally we think that things are normal, and give up this assumption only when some good reason appears for doing so.

These reflections suggest the following *contextualized* restatement of our original definitions of sufficient conditions and necessary conditions:

> That F is a sufficient condition for G means that whenever F is present in a normal context, G is present there as well.
>
> That F is a necessary condition for G means that whenever F is absent from a normal context, G is absent from it as well.

What will count as a normal context will vary with the type and the aim of an investigation, but all investigations into causally sufficient conditions and causally necessary conditions take place against the background of many factors that are taken as fixed.

Background Assumptions. If we are going to subject a causal hypothesis to rigorous testing employing the SCT and the NCT, we have to seek out a wide range of cases in which it might possibly be refuted. In general, the wider the range of possible refuters the better. But some limit must be put on this activity, or else testing will get hopelessly bogged down. If we are testing a drug to see whether it will cure a disease, we should try it on a variety of people of various ages, medical histories, body types, and so on, but we will not check to see whether it works on people named Edmund or check to see whether it works on people who drive Volvos. Such factors, we want to say, are plainly irrelevant. But what makes them irrelevant? How do we distinguish relevant from irrelevant considerations?

The answer to this question is that our reasoning about causes occurs within a framework of beliefs that we consider already established as true. This framework contains a great deal of what is called *common knowledge*—knowledge we expect almost every sane adult to possess. We all know, for example, that human beings cannot breathe underwater, cannot walk through walls, cannot be in two places at once, and so on. The stock of these commonplace beliefs is almost endless. Because they are commonplace beliefs, they tend not to be mentioned, yet they play an important role in distinguishing relevant factors from irrelevant ones. Furthermore, *specialized* knowledge contains its own principles that are largely taken for granted by specialists. Doctors, for example, know a great deal about the detailed structure of the human body, and this background knowledge constantly guides their thought in dealing with specific illnesses. Thus, accepted beliefs establish a framework in which new ideas are judged. We can call the beliefs that give our system of beliefs its basic structure *framework beliefs*. In general, we do not question these framework beliefs, and things that do not fit in with them are not taken seriously. For example, if someone claimed to discover that blood does not circulate, no doctor would take the time to refute her. To take this suggestion seriously would be to challenge our whole way of viewing human beings and many other organisms.

Not all beliefs have this central status in our system of beliefs. They can be abandoned or modified without seriously harming our basic beliefs about the world. Someone might mistakenly believe that Portland is the capital of Oregon. Having discovered that the capital is really Salem, that person will drop one belief and replace it with another. Normally this change in belief will have little effect on many other beliefs and probably no effect on framework beliefs. In contrast, if scientists came to doubt that the sun is larger than the Earth, most of their conception of the physical universe would have to change. This suggests the following picture of our system of beliefs: in it, some beliefs matter more or carry more weight than others, and they are given up only under great pressure. Other beliefs can be given up with hardly a thought. Discovering that Salem, not Portland, is the capital of Oregon can normally be taken in stride; discovering that half the world's population is made up of invading aliens from another planet would profoundly change our way of looking at things.

323

Given this picture, we see that beliefs are not tested in isolation from other beliefs. Testing takes place within a system of beliefs that shapes the form the testing will take. The Austrian philosopher Ludwig Wittgenstein put the matter this way:

> 105. All testing, all confirmation and disconfirmation of a hypothesis takes place already within a system. And this system is not a more or less arbitrary and doubtful point of departure for all our arguments: no, it belongs to the essence of what we call an argument. The system is not so much the point of departure, as the element in which arguments have their life.[15]

We can now see how our belief system develops and grows. Things that have been learned in the past guide the search for new knowledge. They do this in two ways: we use the belief system to reject what we take to be false or irrelevant hypotheses, and we also use it to suggest interesting new hypotheses. New beliefs are added to old beliefs, and these, in turn, suggest further possible new beliefs. The more well-founded beliefs we have, the more we are able to discover further well-founded beliefs. Learning about the world is a bootstrap operation.

The situation, however, is more complicated than this. We not only increase our stock of beliefs but sometimes we discover that some of our previously held beliefs are false. For example, we might learn to our astonishment

[15]Ludwig Wittgenstein, *On Certainty*, trans. G. E. M. Anscombe (Oxford: Basil Blackwell, 1969).

that a close friend is a member of a terrorist organization. Such a discovery almost certainly would lead us to reconsider a great many beliefs we have about him. Occasionally even framework beliefs are called into question and then rejected. The replacement of the Earth-centered (geocentric) theory of the solar system of Ptolemy by the sun-centered (heliocentric) theory of Copernicus is one example of a revolutionary change;[16] the replacement of Newtonian physics by Einstein's relativistic physics is another. But such revolutions in thought are rare. Furthermore, such theories, as novel as they may be, leave most of the general framework beliefs untouched. If Einstein's theory (or any theory) entailed that human beings fly by flapping their ears, that would be enough reason for rejecting it. Even the most revolutionary changes in thought, when looked at from a distance, appear remarkably conservative of our *total* system of beliefs.

A Detailed Example. To get a clearer idea of the complex interplay between the rules we have been examining and the reliance on background information, it will be helpful to look in some detail at actual applications of these rules. For this purpose, we will examine an attempt to find the cause of a particular phenomenon, an outbreak of what came to be known as Legionnaires' disease. The example not only shows how causal reasoning relies on background assumptions, it has another interesting feature as well: in the process of discovering the cause of Legionnaires' disease the investigators were forced to *abandon* what was previously taken to be a well-established causal generalization. In fact, until it was discarded, this false background principle gave them no end of trouble.

The following account of the outbreak of Legionnaires' disease and the subsequent difficulties in finding out what caused it is drawn from an article in *Scientific American*.[17]

> The 58th convention of the American Legion's Pennsylvania Department was held at the Bellevue-Stratford Hotel in Philadelphia from July 21 through 24, 1976. . . . Between July 22 and August 3, 149 of the conventioneers developed what appeared to be the same puzzling illness, characterized by fever, coughing and pneumonia. This, however, was an unusual, explosive outbreak of pneumonia with no apparent cause. . . . Legionnaires' disease, as the illness was quickly named by the press, was to prove a formidable challenge to epidemiologists and laboratory investigators alike.

Notice that at this stage the researchers begin with the assumption that they are dealing with a single illness and not a collection of similar but different illnesses. That assumption could turn out to be wrong; but, if the symptoms of the various patients are sufficiently similar, this is a natural starting assumption. Another reasonable starting assumption is that this illness had a

[16]For the classic discussion of these two systems, see Galileo's *Dialogue Concerning the Two World Systems—Ptolemaic and Copernican*, which is excerpted in Chapter 15.

[17]These excerpts are drawn from David W. Fraser and Joseph E. McDade, "Legionellosis," *Scientific American*, October 1977, 82–99.

single causative agent. This assumption, too, could turn out to be false, though it did not. The assumption that they were dealing with a single disease with a single cause was at least a good simplifying assumption, one to be held onto until there was good reason to give it up. In any case, we now have a clear specification of our target feature, *G*: the occurrence of a carefully described illness that came to be known as Legionnaires' disease. The situation concerning it was puzzling because people had contracted a disease with symptoms much like those of pneumonia, yet they had not tested positive for any of the known agents that cause such diseases.

The narrative continues as follows:

> The initial step in the investigation of any epidemic is to determine the character of the illness, who has become ill and just where and when. The next step is to find out what was unique about the people who became ill: where they were and what they did that was different from other people who stayed well. Knowing such things may indicate how the disease agent was spread and thereby suggest the identity of the agent and where it came from.

Part of this procedure involves a straightforward application of the NCT: was there any interesting feature that was always present in the history of people who came down with the illness? Progress was made almost at once on this front:

324

> We quickly learned that the illness was not confined to Legionnaires. An additional 72 cases were discovered among people who had not been directly associated with the convention. They had one thing in common with the sick conventioneers: for one reason or another they had been in or near the Bellevue-Stratford Hotel.

Strictly speaking, of course, all these people who had contracted the disease had more than one thing in common. They were, for example, all alive at the time they were in Philadelphia, and being alive is, in fact, a necessary condition for getting Legionnaires' disease. But the researchers were not interested in this necessary condition because it is a normal background condition for the contraction of any disease. Furthermore, it did not provide a condition that distinguished those who contracted the disease from those who did not. The overwhelming majority of people who were alive at the time did not contract Legionnaires' disease. Thus, the researchers were not interested in this necessary condition because it would fail so badly when tested as a sufficient condition. On the basis of common knowledge and specialized medical knowledge, a great many other conditions were also kept off the candidate list.

One prime candidate on the list was presence at the Bellevue-Stratford Hotel. The application of the NCT to this candidate was straightforward. Everyone who had contracted the disease had spent time in or near that hotel. Thus, presence at the Bellevue-Stratford could not be eliminated as a necessary condition of Legionnaires' disease.

The application of the SCT was more complicated, because not everyone who stayed at the Bellevue-Stratford contracted the disease. Other factors

made a difference: "Older conventioneers had been affected at a higher rate than younger ones, men at three times the rate for women." Since some young women (among others) who were present at the Bellevue-Stratford did not get Legionnaires' disease, presence at that hotel could be eliminated as a sufficient condition of Legionnaires' disease. Nonetheless, it is part of medical background knowledge that susceptibility to disease often varies with age and gender. Given these differences, some people who spent time at the Bellevue-Stratford were at higher risk for contracting the disease than others. The investigation so far suggested that, for some people, being at the Bellevue-Stratford was connected with a sufficient condition for contracting Legionnaires' disease.

As soon as spending time at the Bellevue-Stratford became the focus of attention, other hypotheses naturally suggested themselves. Food poisoning was a reasonable suggestion, since it is part of medical knowledge that diseases are sometimes spread by food. It was put on the list of possible candidates, but failed. Investigators checked each local restaurant and each function where food and drink were served. Some of the people who ate in each place did not get Legionnaires' disease, so the food at these locations was eliminated by the SCT as a sufficient condition of Legionnaires' disease. These candidates were also eliminated by the NCT as necessary conditions because some people who did get Legionnaires' disease did not eat at each of these restaurants and functions. Thus, the food and drink could not be the cause.

Further investigation turned up another important clue to the cause of the illness.

> Certain observations suggested that the disease might have been spread through the air. Legionnaires who became ill had spent on the average about 60 percent more time in the lobby of the Bellevue-Stratford than those who remained well; the sick Legionnaires' had also spent more time on the sidewalk in front of the hotel than their unaffected fellow conventioneers. . . . It appeared, therefore, that the most likely mode of transmission was airborne.

Merely breathing air in the lobby of the Bellevue-Stratford Hotel still could not be a necessary or sufficient condition, but the investigators reasoned that something in the lobby air probably caused Legionnaires' disease, since the rate of the disease varied up or down in proportion to the time spent in the lobby (or near it on the sidewalk in front). This is an application of the method of Concomitant Variation, which will be discussed below.

Now that the focus was on the lobby air, the next step was to pinpoint a specific cause in that air. Again appealing to background medical knowledge, there seemed to be three main candidates for the airborne agents that could have caused the illness: "heavy metals, toxic organic substances, and infectious organisms." However, examination of tissues taken from patients who had died from the disease revealed "no unusual levels of metallic or toxic organic substances that might be related to the epidemic," so this left an infectious organism as the remaining candidate. Once more we have an

325

application of NCT. If the disease had been caused by heavy metals or toxic organic substances, then there would have been unusually high levels of these substances in the tissues of those who had contracted the disease. Because this was not so, these candidates were eliminated as necessary conditions of the disease.

Appealing to background knowledge once more, it seemed that a bacterium would be the most likely source of an airborne disease with the symptoms of Legionnaires' disease. But researchers had already made a routine check for bacteria that cause pneumonialike diseases, and they had found none. For this reason, attention was directed to the possibility that some previously unknown organism had been responsible but had somehow escaped detection.

It turned out that an undetected and previously unknown bacterium *had* caused the illness, but it took more than four months to find this out. The difficulties encountered in this effort show another important fact about the reliance on a background assumption: sometimes it turns out to be false. To simplify, the standard way to test for the presence of bacteria is to try to grow them in culture dishes—flat dishes containing nutrients that bacteria can live on. If, after a reasonable number of tries, a colony of a particular kind of bacterium does not appear, then it is concluded that the bacterium is not present. As it turned out, the bacterium that caused Legionnaires' disease would not grow in the cultures commonly used to detect the presence of bacteria. Thus, an important background assumption turned out to be false.

After a great deal of work, a suspicious bacterium was detected using a live-tissue culture rather than the standard synthetic culture. The task, then, was to show that this particular bacterium in fact caused the disease. Again to simplify, when people are infected by a particular organism, they often develop antibodies that are specifically aimed at this organism. In the case of Legionnaires' disease, these antibodies were easier to detect than the bacterium itself. They also remained in the patients' bodies after the infection had run its course. We thus have another chance to apply the NCT: if Legionnaires' disease was caused by this particular bacterium, then whenever the disease was present, this antibody should be present as well. The suspicious bacterium passed this test with flying colors and was named, appropriately enough, *Legionella pneumophila*. Because the investigators had worked so hard to test such a wide variety of candidates, they assumed that the disease must have some cause among the candidates that they checked; so, since only one candidate remained, they felt justified in reaching a positive conclusion that the bacterium was a necessary condition of Legionnaires' disease.

The story of the search for the cause of Legionnaires' disease brings out two important features of the use of inductive methods in the sciences. First, it involves a complicated interplay between what is already established and what is being tested. Confronted with a new problem, established principles can be used to suggest theoretically significant hypotheses to be tested. The tests then eliminate some hypotheses and leave others. If at the end of the

investigation a survivor remains that fits in well with our previously established principles, then the stock of established principles is increased. The second thing that this example shows is that the inductive method is fallible. Without the background of established principles, the application of inductive principles like the NCT and the SCT would be undirected; yet sometimes these established principles let us down, for they can turn out to be false. The discovery of the false background principle that hindered the search for the cause of Legionnaires' disease led to important revisions in laboratory techniques. The discovery that fundamental background principles are false can lead to revolutionary changes in science. This topic is discussed in Chapter 15, where the nature of scientific frameworks is examined.

Calling Things Causes. After their research was finally completed, with the bacterium identified, described, and named, it was then said that *Legionella pneumophila* was the *cause* of Legionnaires' disease. What was meant by this? To simplify a bit, suppose *L. pneumophila* (as it is abbreviated) entered the bodies of *all* those who contracted the disease: whenever the disease was present, *L. pneumophila* was present. Thus, *L. pneumophila* passes the NCT. We will further suppose, as is common in bacterial infections, that some people's immune systems were successful in combating *L. pneumophila*, and they never actually developed the disease. Thus, the presence of *L. pneumophila* would not pass the SCT. This suggests that we sometimes call something a cause if it passes the NCT, even if it does not pass the SCT.

But even if we sometimes consider necessary conditions to be causes, we certainly do not consider *all* necessary conditions to be causes. We have already noted that to get Legionnaires' disease, one has to be alive, yet no one thinks that being alive is the cause of Legionnaires' disease. To cite another example, this time one that is not silly, it might be that another necessary condition for developing Legionnaires' disease is that the person be in a run-down condition—healthy people might always be able to resist *L. pneumophila*. Do we then want to say that being in a run-down condition is the cause of Legionnaires' disease? As we have described the situation, almost certainly not, but we might want to say that it is an important *causal factor* or *causally relevant factor*.

Although the matter is far from clear, what we call *the* cause rather than simply *a* causal factor or causally relevant factor seems to depend on a number of considerations. We tend to reserve the expression "the cause" for *changes* that occur prior to the effect, and describe *permanent* or *standing* features of the context as causal factors instead. That is how we speak about Legionnaires' disease. Being exposed to *L. pneumophila*, which was a specific event that occurred before the onset of the disease, *caused it*. Being in a run-down condition, which was a feature that patients possessed for some time before they contracted the disease, was not called the cause, but instead called a causal factor. It is not clear, however, that we always draw the distinction between what we call the cause and what we call a causal factor based on whether something is a prior event or a standing condition. For example, if we are trying to explain why certain people who came in contact with *L. pneumophila* contracted the disease whereas others did not, then we might say that the former group contracted the disease because they were in a run-down condition. Thus, by limiting our investigation only to those who came in contact with *L. pneumophila*, our perspective has changed. We want to know why some within that group contracted the disease and others did not. Citing the run-down condition of those who contracted the disease as the cause now seems entirely natural. These examples suggest that we call something *the* cause when it plays a particularly important role relative to the purposes of our investigation. Usually this will be an event or change taking place against the background of fixed necessary conditions; sometimes it will not.

Sometimes we call *sufficient* conditions causes. We say that short circuits cause fires because in many normal contexts a short circuit is sufficient to cause a fire. Of course, short circuits are not necessary to cause a fire, because, in the same normal contexts, fires can be caused by a great many other things. With sufficient conditions, as with necessary conditions, we often draw a distinction between what we call the cause as opposed to what we call a causal factor, and we seem to draw it along similar lines. Speaking loosely, we might say that we sometimes call the *key* components of sufficient conditions *causes*. Then, holding background conditions fixed, we can use the SCT to evaluate such causal claims.

In sum, we can use the NCT to eliminate proposed necessary causal conditions. We can use the SCT to eliminate proposed sufficient causal conditions. Those candidates that survive these tests may be called causal conditions or causal factors if they fit in well with our system of other causal generalizations. Finally, some of these causal conditions or causal factors will be called causes if they play a key role in our causal investigations. Typically, though not always, we call something the cause of an event if it is a prior event or change that stands out against the background of fixed conditions.

CONCOMITANT VARIATION

The use of the sufficient condition test and the necessary condition test depends on certain features of the world being sometimes present and sometimes absent. However, certain features of the world are always present to some degree. Because they are always present, the NCT will never eliminate them as possible necessary conditions of any event, and the SCT will never eliminate anything as a sufficient condition for them. Yet the *extent* or *degree* to which a feature exists in the world is often a significant phenomenon that demands causal explanation.

An example should make this clear. In recent years, a controversy has raged over the impact of acid rain on the environment of the northeastern United States and Canada. Part of the controversy involves the proper interpretation of the data that have been collected. The controversy has arisen for

the following reason: the atmosphere always contains a certain amount of acid, much of it from natural sources. It is also known that an excess of acid in the environment can have severe effects on both plants and animals. Lakes are particularly vulnerable to the effects of acid rain. Finally, it is also acknowledged that industries, mostly in the Midwest, discharge large quantities of sulfur dioxide (SO_2) into the air, and this increases the acidity of water in the atmosphere. The question—and here the controversy begins—is whether the contribution of acid from these industries is the cause of the environmental damage downwind of them.

How can we settle such a dispute? The two rules we have introduced provide no immediate help, for, as we have seen, they provide a rigorous test of a causal hypothesis only when we can find contrasting cases with the presence or the absence of a given feature. The NCT provides a rigorous test for a necessary condition only if we can find cases in which the feature does not occur and then check to make sure that the target feature does not occur either. The SCT provides a rigorous test for a sufficient condition only when we can find cases in which the target phenomenon is absent and then check whether the candidate sufficient condition is absent as well. In this case, however, neither check applies, for there is always a certain amount of acid in the atmosphere, so it is not possible to check what happens when atmospheric acid is completely absent. Similarly, environmental damage, which is the target phenomenon under investigation, is so widespread in our modern industrial society that it is also hard to find a case in which it is completely absent.

So, if there is always acid in the atmosphere, and environmental damage always exists at least to some extent, how can we determine whether the SO_2 released into the atmosphere is *significantly* responsible for the environmental damage in the affected areas? Here we use what John Stuart Mill called the *Method of Concomitant Variation*. We ask whether the amount of environmental damage varies directly in proportion to the amount of SO_2 released into the environment. If environmental damage increases with the amount of SO_2 released into the environment and drops when the amount of SO_2 is lowered, this means that the level of SO_2 in the atmosphere is *positively correlated* with environmental damage. We would then have good reason to believe that lowering SO_2 emissions would lower the level of environmental damage, at least to some extent.

Arguments relying on the method of concomitant variation are difficult to evaluate, especially when there is no generally accepted background theory that makes sense of the concomitant variation. Some such variations are well understood. For example, everyone knows that the faster you drive, the more gasoline you consume. (Gasoline consumption varies *directly* with speed.) Why? There is a good theory here; it takes more energy to drive at a high speed than at a low speed, and this energy is derived from the gasoline consumed in the car's engine. Other correlations are less well understood. There seems to be a correlation between cholesterol level in the blood and the chances of heart attack. First, the correlation here is not nearly so good as the gasoline-consumption-to-speed correlation, for many people with high cholesterol levels do not suffer heart attacks, and many people with low cholesterol levels do. Furthermore, no generally accepted background theory has been found that explains the positive correlation that does seem to exist.

This reference to background theory is important, because two sets of phenomena can be correlated to a very high degree, even with no direct causal relationship between them. A favorite example that appears in many statistics texts is the discovered positive correlation in boys between foot size and quality of handwriting. It is hard to imagine a causal relation holding in either direction. Having big feet should not make you write better and, just as obviously, writing well should not give you big feet. The correct explanation is that both foot size and handwriting ability are positively correlated with age. Here a noncausal correlation between two phenomena (foot size and handwriting ability) is explained by a third common correlation (maturation) that *is* causal.

At times, it is possible to get causal correlations *backward*. For example, a few years ago, sports statisticians discovered a negative correlation between forward passes thrown and winning. That is, the more forward passes a team threw, the less chance it had of winning. This suggests that passing is not a good strategy, since the more you do it, the more likely you are to lose. Closer examination showed, however, that the causal relationship, in fact, went in the other direction. Toward the end of a game, losing teams tend to throw a great many passes in an effort to catch up. In other words, teams throw a lot of passes because they are losing, rather than the other way around.

Finally, some correlations seem inexplicable. For example, a strong positive correlation holds between the birth rate in Holland and the number of storks nesting in chimneys. There is, of course, a background theory that would explain this—storks bring babies—but that theory is not favored by modern science. For the lack of any better background theory, the phenomenon just seems weird.

So, given a strong correlation between phenomena of types *A* and *B*, four possibilities exist:

(1) *A* is the cause of *B*.
(2) *B* is the cause of *A*.
(3) Some third thing is the cause of both.
(4) The correlation is simply accidental.

Before we accept any one of these possibilities, we must have good reasons for preferring it over the other three.

327

EXERCISE XII

In each of the following examples a strong correlation, either negative or positive, holds between two sets of phenomena, *A* and *B*. Try to decide whether *A* is the cause of *B*, *B* is the cause of *A*, both are caused by some third factor, C, or the correlation is simply accidental. Explain your choice.

(1) At one time there was a strong negative correlation between the number of mules in a state (*A*) and the salaries paid to professors at the state university (*B*). In other words, the more mules, the lower professional salaries.[18]

(2) It has been claimed that there is a strong positive correlation between those students who take sex education courses (*A*) and those who contract venereal disease (*B*).

(3) LOCKED DOORS NO BAR TO CRIME, STUDY SAYS

"Washington (UPI)—Rural Americans with locked doors, watchdogs or guns may face as much risk of burglary as neighbors who leave doors unlocked, a federally financed study says.

"The study, financed in part by a three-year $170,000 grant from the Law Enforcement Assistance Administration, was based on a survey of nearly 900 families in rural Ohio.

"Sixty percent of the rural residents surveyed regularly locked doors [*A*], but were burglarized more often than residents who left doors unlocked [*B*]."[19]

(4) There is a high positive correlation between the number of fire engines in a particular borough in New York City (*A*) and the number of fires that occur there (*B*).[20]

(5) For a particular United States president, there is a negative correlation between the number of hairs on his head (*A*) and the population of China (*B*).

DISCUSSION QUESTIONS

(1) Many defenders of nuclear deterrence have relied on an inductive argument to the effect that World War III was avoided because of the balance of power between West and East. What evidence has been offered in support of this conclusion? How strong is the argument?

(2) Now that it seems beyond doubt that smoking is dangerous to people's health, a new debate has arisen concerning the possible health hazards of smoke on nonsmokers. Collect statements pro and con on this issue and evaluate the strength of the inductive arguments on each side.

(3) In *Twilight of the Idols*,[21] Nietzsche claims that the following examples illustrate "the error of mistaking cause for consequence." Do you agree? Why or why not?

[21]Friedrich Nietzsche, *Twilight of the Idols* and *The Anti-Christ*, trans. R. J. Hollingdale (1889; Harmondsworth: Penguin, 1968).

[18]From Gregory A. Kimble, *How to Use (and Misuse) Statistics* (Englewood Cliffs, NJ: Prentice-Hall, 1978), 182.

[19]"Locked Doors No Bar to Crime, Study Says," *Santa Barbara* [California] *Newspress*, Wednesday, February 16, 1977.

[20]From Kimble, *How to Use (and Misuse) Statistics*, 182.

> Everyone knows the book of the celebrated Cornaro in which he recommends his meagre diet as a recipe for a long and happy life—a virtuous one, too. Few books have been so widely read; even now many thousands of copies are printed in England every year. I do not doubt that hardly any book (the Bible rightly excepted) has done so much harm, has shortened so many lives, as this curiosity, which was so well meant. The reason: mistaking the consequence for the cause. The worthy Italian saw in his diet the *cause* of his long life: while the prerequisite of long life, an extraordinarily slow metabolism, a small consumption, was the cause of his meagre diet. He was not free to eat much *or* little as he chose, his frugality was *not* an act of "free will": he became ill when he ate more. But if one is not a bony fellow of this sort one does not merely do well, one positively needs to eat *properly*. A scholar of *our* day, with his rapid consumption of nervous energy, would kill himself with Cornaro's regimen. . . .
>
> Long life, a plentiful posterity is *not* the reward of virtue, virtue itself is rather just that slowing down of the metabolism which also has, among other things, a long life, a plentiful posterity, in short *Cornarism*, as its outcome.—The Church and morality say: "A race, a people perishes through vice and luxury". My *restored* reason says: when a people is perishing, degenerating physiologically, vice and luxury (that is to say the necessity for stronger and stronger and more and more frequent stimulants, such as every exhausted nature is acquainted with) *follow* therefrom. A young man grows prematurely pale and faded. His friends say: this and that illness is to blame. I say: *that* he became ill, *that* he failed to resist the illness, was already the consequence of an impoverished life, an hereditary exhaustion. The newspaper reader says: this party will ruin itself if it makes errors like this. My *higher* politics says: a party which makes errors like this is already finished—it is no longer secure in its instincts. Every error, of whatever kind, is a consequence of degeneration of instinct, disgregation of will: one has thereby virtually defined the *bad*. Everything *good* is instinct—and consequently easy, necessary, free. Effort is an objection, the *god* is typically distinguished from the hero (in my language: *light* feet are the first attribute of divinity).

INDUCTIVE GENERALIZATION

To apply the method of concomitant variation to the issue of whether smoking causes lung cancer, we need to know the rates of smoking and the rates of lung cancer. In this case and many others like it, the crucial relationships do not hold universally, so we cannot apply the tests for necessary and sufficient conditions that were discussed above. How, then, can we determine those rates?

The answer lies in another common form of inductive reasoning, which is often called *inductive generalization*. In inductive generalizations, we cite characteristics of a sample of a population to support a claim about the character of the population as a whole. Opinion polls work this way. Suppose a candidate wants to know how popular she is with voters. Because it would be practically impossible to survey all voters, she takes a sample of voting opinion and then infers that the opinions of those sampled indicate the overall opinion of voters. Thus, if 60 percent of the voters sampled say that they will vote for her, she concludes that she will get more or less 60 percent of the vote in the actual election. As we shall see later, inferences of this kind often go wrong, even when made by experts, but the general pattern of this reasoning is quite clear: statistical features of a sample are used to make statistical claims about the population as a whole.

Basically the same form of reasoning can be used to reach a universal conclusion. An example is the inductive inference discussed at the start of this chapter: all observed ravens are black, so all ravens are black. Again, we sample part of a population to draw a conclusion about the whole.

How do we assess such inferences? To begin to answer this question, we can consider a simple example of an inductive generalization. On various occasions Harold has tried to use Canadian quarters in American telephones and found that they have not worked. From this he draws the conclusion that Canadian quarters do not work in American telephones. Harold's inductive reasoning looks like this:

> In the past, when I tried to use Canadian quarters in American telephones, they have not worked.
>
> ∴ Canadian quarters do not work in American telephones.

The force of the conclusion is that Canadian quarters *never* work in American telephones.

In evaluating this argument, what questions should we ask? We can start with a question that we should ask of any argument.

SHOULD WE ACCEPT THE PREMISES?

Perhaps Harold has a bad memory, has kept bad records, or is a poor observer. For some obscure reason, he may even be lying. It is important to ask this question explicitly, because fairly often the premises, when challenged, will not stand up to scrutiny.

If we decide that the premises are acceptable (that is, true and justified), then we can shift our attention to the relationship between the premises and the conclusion and ask how much support the premises give to the conclusion. One commonsense question is this: "Just how many times has Harold tried to use Canadian quarters in American telephones?" If the answer is "Once," then our confidence in his argument should drop to almost nothing. So, for inductive generalizations, it is always appropriate to ask about the size of the sample.

IS THE SAMPLE LARGE ENOUGH?

One reason we should be suspicious of small samples is that they can be affected by runs of luck. Suppose Harold flips a Canadian quarter four times and it comes up heads each time. From this, he can hardly conclude that Canadian quarters always come up heads when flipped. He could not even reasonably conclude that *this* Canadian quarter would always come up heads when flipped. The reason for this is obvious enough; if you spend a lot of time flipping coins, runs of four heads in a row are not all that unlikely (the probability is actually one in sixteen), and therefore samples of this size can easily be distorted by chance. On the other hand, if Harold flipped the coin twenty times and it continued to come up heads, he would have strong grounds for saying that this coin, at least, will always come up heads. In fact, he would have strong grounds for thinking that he has a two-headed coin. Because an overly small sample can lead to erroneous conclusions, we need to make sure that our sample includes enough trials.

How many is enough? On the assumption, for the moment, that our sampling has been fair in all other respects, how many samples do we need to provide the basis for a strong inductive argument? This is not always an easy question to answer, and sometimes answering it demands subtle mathematical techniques. Suppose your company is selling 10 million computer chips to the Department of Defense, and you have guaranteed that no more than 0.2 percent of them will be defective. It would be prohibitively expensive to test all the chips, and testing only a dozen would hardly be enough to reasonably guarantee that the total shipment of chips meets the required specifications. Because testing chips is expensive, you want to test as few as possible, but because meeting the specifications is crucial, you want to test enough to guarantee that you have done so. Answering questions of this kind demands sophisticated statistical techniques beyond the scope of this text.

329

Sometimes, then, it is difficult to decide how many samples are needed to give reasonable support to inductive generalizations; yet many times it is obvious, without going into technical details, that the sample is too small. Drawing an inductive conclusion from a sample that is too small can lead to the fallacy of *hasty generalization*. It is surprising how common this fallacy is. We see a person two or three times and find him cheerful, and we immediately leap to the conclusion that he is a cheerful person. That is, from a few instances of cheerful behavior, we draw a general conclusion about his personality. When we meet him later and find him sad, morose, or grouchy, we then conclude that he has changed—thus swapping one hasty generalization for another.

This tendency toward hasty generalization was discussed more than 200 years ago by the philosopher David Hume, who saw that we have a strong tendency to "follow general rules which we rashly form to ourselves, and which are the source of what we properly call prejudice."[22] This tendency toward hasty generalization has been the subject of extensive psychological investigation. The cognitive psychologists Amos Tversky and Daniel Kahneman put the matter this way:

> We submit that people view a sample randomly drawn from a population as highly representative, that is, similar to the population in all essential characteristics. Consequently, they expect any two samples drawn from a particular population to be more similar to one another and to the population than sampling theory predicts, at least for small samples.[23]

[22]David Hume, *A Treatise of Human Nature*, 2d ed. (1739; Oxford: Oxford University Press, 1978), 146.

To return to a previous example, we make our judgments of someone's personality on the basis of a very small sample of his or her behavior and expect this person to behave in similar ways in the future when we encounter further samples of behavior. We are surprised, and sometimes indignant, when the future behavior does not match our expectations.

By making our samples sufficiently large, we can guard against distortions due to "runs of luck," but even very large samples can give us a poor basis for an inductive generalization. Suppose that Harold has tried hundreds of times to use a Canadian quarter in an American telephone, and it has never worked. This will increase our confidence in his inductive generalization, but size of sample alone is not a sufficient ground for a strong inductive argument. Suppose that Harold has tried the same coin in hundreds of different telephones, or tried a hundred different Canadian coins in the same telephone. In the first case, there might be something wrong with this particular coin; in the second case, there might be something wrong with this particular telephone. In neither case would he have good grounds for making the general claim that *no* Canadian quarters work in *any* American telephones. This leads us to the third question we should ask of any inductive generalization.

Is the Sample Biased?

When the sample, however large, is not representative of the population, then it is said to be unfair or biased. Here we can speak of the fallacy of *biased sampling*. One of the most famous errors of biased sampling was committed by a magazine named the *Literary Digest*. Before the presidential election of 1936, this magazine sent out 10 million questionnaires asking which candidate the recipient would vote for: Franklin Roosevelt or Alf Landon. It received 2.5 million returns, and on the basis of the results confidently predicted that Landon would win by a landslide: 56 percent for Landon to only 44 percent for Roosevelt. When the election results came in, Roosevelt had won by an even larger landslide in the opposite direction: 62 percent for Roosevelt to a mere 38 percent for Landon. What went wrong? The sample was certainly large enough; in fact, by contemporary standards it was much larger than needed. It was the way the sample was selected, not its size, that caused the problem: the sample was randomly drawn from names in telephone books and from club membership lists. In 1936 there were only 11 million telephones in the United States, and many of the poor—especially the rural poor—did not have telephones. During the Great Depression there were more than 9 million unemployed in America; they were almost all poor and thus underrepresented on club membership lists. Finally, a large

[23]Amos Tversky and Daniel Kahneman, "Belief in the Law of Small Numbers," *Psychological Bulletin* 76, no. 2 (1971): 105.

percentage of these underrepresented groups voted for Roosevelt, the Democratic candidate. As a result of these biases in its sampling, along with some others, the *Literary Digest* underestimated Roosevelt's percentage of the vote by a whopping 18 percent.

Looking back, it may be hard to believe that intelligent observers could have done such a ridiculously bad job of sampling opinion, but the story repeats itself, though rarely on the grand scale of the *Literary Digest* fiasco. In 1948, for example, the Gallup poll, which had correctly predicted Roosevelt's victory in 1936, predicted, as did other major polls, a clear victory for Thomas Dewey over Harry Truman. Confidence was so high in this prediction that the *Chicago Tribune* published a banner headline declaring that Dewey had won the election before the votes were actually counted. What went wrong this time? The answer here is more subtle. The Gallup pollsters (and others) went to great pains to make sure that their sample was representative of the voting population. The interviewers were told to poll a certain number of people from particular social groups—rural poor, suburban middle class, urban middle class, ethnic minorities, and so on—so that the proportions of those interviewed matched, as closely as possible, the proportions of those likely to vote. (The *Literary Digest* went bankrupt after its incorrect prediction, so the pollsters were taking no chances.) Yet somehow bias crept into the sampling; the question was, how? One speculation was that a large percentage of those sampled did not tell the truth when they were interviewed; another was that a large number of people changed their minds at the last minute. So perhaps the data collected were not reliable. The explanation generally accepted was more subtle. Although Gallup's workers were told to interview specific numbers of people from particular classes (so many from the suburbs, for example), they were not instructed to choose people randomly from within each group. Without seriously thinking about it, they tended to go to "nicer" neighborhoods and interview "nicer" people. Because of this, they biased the sample in the direction of their own (largely) middle-class preferences and, as a result, underrepresented constituencies that would give Truman his unexpected victory.

38

Is the Result Biased in Some Other Way?

Because professionals using modern statistical analysis can make bad inductive generalizations through biased sampling, it is not surprising that our everyday, informal inductive generalizations are often inaccurate. Sometimes we go astray because of small samples and biased samples. This happens, for example, when we form opinions about what people think or what people are like by asking only our friends. But bias can affect our reasoning in other ways as well.

One of the main sources of bias in everyday life is *prejudice*. Even if we sample a wide enough range of cases, we often reinterpret what we hear or see in light of some preconception. People who are prejudiced will find very little good and a great deal bad in those they despise, no matter how these people actually behave. In fact, most people are a mixture of good and bad qualities. By ignoring the former and dwelling on the latter, it is easy enough for a prejudiced person to confirm negative opinions. Similarly, stereotypes, which can be either positive or negative, often persist in the face of overwhelming counterevidence. Criticizing the beliefs common in Britain in his own day, David Hume remarked:

> An Irishman cannot have wit, and a Frenchman cannot have solidity; for which reason, though the conversation of the former in any instance be very agreeable, and of the latter very judicious, we have entertained such a prejudice against them, that they must be dunces and fops in spite of sense and reason.[24]

Although common stereotypes have changed somewhat since Hume's day, prejudice continues to distort the beliefs of many people in our own time.

Another common source of bias in sampling arises from phrasing questions in ways that encourage certain answers while discouraging others. Even if a fair sample is asked a question, it is well known that the way a question is phrased can exert a significant influence on how people will answer it. Questions like the following are not intended to illicit information, but instead to push people's answers in one direction rather than another.

(1) Which do you favor: (a) preserving a citizen's constitutional right to bear arms or (b) leaving honest citizens defenseless against armed criminals?

(2) Which do you favor: (a) restricting the sale of assault weapons or (b) knuckling under to the demands of the well-financed gun lobby?

In both cases one alternative is made to sound attractive, the other unattractive. When questions of this sort are used, it is not surprising that different pollsters can come up with wildly different results.

Summary

Now we can summarize and restate our questions with more accuracy. Confronted with inductive generalizations, there are four questions that we should routinely ask:

(1) Are the premises acceptable?

(2) Is the sample too small?

(3) Is the sample biased?

(4) Are the results affected by other sources of bias?

Exercise XIII

By asking the preceding questions, specify what, if anything, is wrong with the following inductive generalizations:

[24] Hume, *A Treatise of Human Nature*, 146–47.

(1) K-Mart asked all of their customers throughout the country whether they prefer K-Mart to Walmart, and 90 percent said they did, so 90 percent of all shoppers in the area prefer K-Mart.

(2) A Swede stole my bicycle, so most Swedes are thieves.

(3) I've never tried it before, but I just put a kiwi fruit in a tub of water. It floated, so most kiwi fruits float in water.

(4) Mary told me that all of her older children are geniuses, so her baby will probably be a genius, too.

(5) When asked whether they would prefer a tax break or a bloated budget, almost everyone said that they wanted a tax break. So a tax break is overwhelmingly popular with the people.

(6) When hundreds of convicted murderers in states without the death penalty were asked whether they would have committed the murder if the state had a death penalty, most of them said that they would not have done it. So most murders can be deterred by the death penalty.

EXERCISE XIV

331

Ann Landers caused a stir when she announced the results of a mail poll that asked her women readers to respond to the following question:

Would you be content to be held and treated tenderly, and forget about "the act"?

Her readers were instructed to answer "Yes" or "No" and indicate whether they were over (or under) forty years of age.

The result was that 72 percent of the respondents answered "Yes," and of those who answered "Yes," 40 percent indicated that they were under forty years of age.

What are we to make of these results? Ann Landers expressed surprise that so many of those who answered "Yes" came from the under-forty group. But for her, "the greatest revelation" was "what the poll says about men as lovers. Clearly, there is trouble in paradise" (*Ask Ann Landers*, January 14 and 15, 1985).

The poll, of course, did not employ scientific methods of sampling (nor did Ann Landers claim that it did), so it is important to look for sources of bias before drawing any conclusions from this poll. Discuss at least three possible sources of bias that could make the Landers sample unrepresentative of the opinions of the population of adult women in America.

STATISTICAL SYLLOGISMS

In a statistical generalization we draw inferences concerning a population from information concerning a sample of that population. From the fact that 60 percent of the population sampled said that they would vote for candidate *X*, we might draw the conclusion that roughly 60 percent of the population will vote for candidate *X*. With a *statistical syllogism* we reason in the reverse

direction: from information concerning a population, we draw a conclusion concerning a member or subset of that population. Here is an example:

> Ninety-seven percent of the Republicans in California voted for Bush.
> Marvin is a Republican from California.
>
> ∴ Marvin voted for Bush.

Such arguments have the following general form:

> *X* percent of *F*s have the feature *G*.
> *a* is an *F*.
>
> ∴ *a* has the feature *G*.[25]

Obviously, when we evaluate the strength of a statistical syllogism, the percentage of *F*s that have the feature *G* will be important. As the figure approaches 100 percent, the statistical argument gains strength. Thus our original argument concerning Marvin is quite strong. We can also get strong statistical syllogisms when the figure approaches 0 percent. The following is a strong inductive argument:

> Three percent of the socialists in California voted for Bush.
> Maureen is a socialist from California.
>
> ∴ Maureen did *not* vote for Bush.

Statistical syllogisms of the kind considered here will be strong only if the figures are close to 100 percent or 0 percent. When the percentages are in the middle of this range, statistical syllogisms are weak.

A more interesting problem in evaluating the strength of a statistical syllogism concerns the *relevance* of the premises to the conclusion. In the above schematic representation of a statistical syllogism, *F* stands for what is called the *reference class*. In our first example, being a Republican from California is the reference class; in our second example, being a socialist from California is the reference class. A striking feature of statistical syllogisms is that using different reference classes can yield incompatible results. To see this, consider the following statistical syllogism:

> Three percent of Gore's relatives voted for Bush.
> Marvin is a relative of Gore.
>
> ∴ Marvin did not vote for Bush.

We now have a statistical syllogism that gives us strong support for the claim that Marvin did not vote for Bush, and this is incompatible with our first

[25]We can also have a *probabilistic* version of the statistical syllogism:

> Ninety-seven percent of the Republicans in California voted for Bush.
> Marvin is a Republican from California.
>
> ∴ There is a 97 percent chance that Marvin voted for Bush.

We will discuss arguments concerning probability in the next chapter.

statistical syllogism, which gave strong support to the claim that he did. To overlook this conflict between statistical syllogisms based on different reference classes would be a kind of fallacy. Which statistical syllogism, if either, should we trust? This will depend on which of the reference classes we take to be more relevant. Which counts more, political affiliation or family ties? That might be hard to say.

One way of dealing with competing statistical syllogisms is to combine the reference classes. We could ask, for example, what percentage of Republicans from California who are relatives of Gore voted for Bush? The result might come out this way:

> Forty-two percent of Republicans from California who were relatives of Gore voted for Bush.
> Marvin is a Republican from California who is a relative of Gore.
> ∴ Marvin voted for Bush.

This statistical syllogism provides very weak support for its conclusion. It supplies stronger, but still weak, support for the denial of the conclusion—that is, that Marvin did not vote for Bush.

This series of arguments illustrates in a clear way what we earlier called the defeasibility of inductive inferences: a strong inductive argument can be made weak by adding further information to the premises. Given that Marvin is a Republican from California, we seemed to have good reason to think that he voted for Bush. But when we added to this the additional piece of information that he was a relative of Gore, the original argument lost most of its force. And new information could produce another reversal. Suppose we discover that Marvin, though a relative of Gore, actively campaigned for Bush. Just about everyone who actively campaigns for a candidate votes for that candidate, so it seems that we again have good reason for thinking that Marvin voted for Bush.

332

It is clear, then, that the way we select our reference classes will affect the strength of a statistical syllogism. The general idea is that we should define our reference classes in a way that brings all relevant evidence to bear on the subject. But this raises difficulties. It is not always obvious which factors are relevant and which factors are not. In our example, party affiliation is relevant to how people voted in the 2000 election; shoe size presumably is not. Whether gender is significant, and, if so, how significant, is a matter for further statistical research.

These difficulties concerning the proper way to fix reference classes reflect a feature of all inductive reasoning: to be successful, such reasoning must take place within a broader framework that helps determine which features are significant and which features are not. Without this framework, there would be no reason not to consider shoe size when trying to decide how someone will vote. This shows how statistical syllogisms, like all of the other inductive arguments that we studied in this chapter, cannot work properly without appropriate background assumptions.

EXERCISE XV

Carry the story of Marvin two steps further, producing two more reversals in the strength of the statistical syllogism with the conclusion that Marvin voted for Bush.

EXERCISE XVI

For each of the following statistical syllogisms, identify the reference class, and then evaluate the strength of the argument in terms of the percentages or proportions cited and the relevance of the reference class.

(1) Less than 1 percent of the people in the world voted for Bush.
Gale is a person in the world.
∴ Gale did not vote for Bush.

(2) Very few teams repeat as Super Bowl champions.
San Francisco was the last Super Bowl champion.
∴ San Francisco will not repeat as Super Bowl champion.

(3) A very high percentage of people in the Senate are men.
Nancy Katzenbaum is in the Senate.
∴ Nancy Katzenbaum is a man.

(4) Three percent of socialists with blue eyes voted for Bush.
Maureen is a socialist with blue eyes.
∴ Maureen did not vote for Bush.

(5) Ninety-eight percent of what John says is true.
John said that the Giants are going to win.
∴ The Giants are going to win.

(6) Half the time he doesn't know what he is doing.
He is eating lunch.
∴ He does not know he is eating lunch.

DISCUSSION QUESTION

Although both in science and in daily life, we rely heavily on the methods of inductive reasoning, a number of perplexing problems exist concerning the legitimacy of this kind of reasoning. The most famous problem concerning induction was formulated by the eighteenth-century philosopher David Hume, first in his *Treatise of Human Nature* and then later in his *Enquiry Concerning Human Understanding*. A simplified version of Hume's skeptical argument goes as follows: our inductive generalizations seem to rest on the assumption that *unobserved* cases will follow the patterns that we discovered in *observed* cases. That is, our inductive generalizations seem to presuppose that nature operates uniformly: the way things are observed to behave here and now are accurate indicators of how things behave anywhere and at any time. But by what right can we assume that nature is uniform? Because this claim itself asserts a contingent matter of fact, it could only be established by inductive reasoning. But because all inductive reasoning presupposes the principle that nature is uniform, any inductive justification of this principle would seem to be circular. It seems, then, that we have no ultimate justification for our inductive reasoning at all. Is this a good argument or a bad one?

333

10

TAKING CHANCES

This chapter offers an elementary discussion of probability and decision making. After showing how informal heuristics *can lead to confusions about probability, formal* laws of probability *are presented along with* Bayes's theorem. *These tools are then used to explain the related notions of* expected monetary value *and* overall value, *which help us assess reasoning about choices involving* risk. *Decisions under* ignorance *pose separate problems, for which a number of rules have been proposed. After discussing some of these rules, the chapter concludes with an examination of two common mistakes in reasoning about probabilities: committing the so-called* gambler's fallacy *and failing to understand the phenomenon of* regression to the mean.

HEURISTICS

Suppose that you are in the Los Angeles airport ready to board a plane to fly to your sister's wedding today outside of Detroit when your flight is suddenly canceled. The airline's agent tells you that you can either book a seat on the next direct flight to Detroit, which is scheduled to leave in three hours, or you can board a flight right now for Chicago and then go stand-by on a flight from Chicago to Detroit. What should you do? That depends on how likely it is that a seat will be available on the flight from Chicago to Detroit. It also depends on how likely it is that you will make it to the wedding in time if you take the direct flight to Detroit three hours late.

In similar ways, many of our most important decisions are based on judgments about probability. Mistakes about probability can then lead to disaster. Doctors lose patients' lives, stockbrokers lose clients' money, and coaches lose games when they underestimate or overestimate probabilities. To avoid such mistakes and losses, we need some way to determine probabilities accurately.

Unfortunately, the most accurate methods for assessing probability are usually not available. In daily life, we have to make a great many decisions, some of them important, most of them not. These decisions often have to be made quickly without pausing to weigh evidence carefully. To deal with this