

# Causal Reasoning

- 12.1 Cause and Effect
- 12.2 Causal Laws and the Uniformity of Nature
- 12.3 Induction by Simple Enumeration
- 12.4 Methods of Causal Analysis
- 12.5 Limitations of Inductive Techniques

## 12.1 Cause and Effect

---

Induction goes far beyond analogical arguments. When we know, or think we know, that one thing is the *cause* of another, or the *effect* of another, we can reason from cause to effect, or from effect to cause. If the supposed relations between cause and effect have been correctly established, the reasoning based on those relations is very powerful.

**Causal reasoning** is also of the very greatest practical importance. Our ability to control our environment, to live successfully and to achieve our purposes, depends critically on our knowledge of causal connections. To cure some disease, for example, physicians must know its cause—and of course they must learn the effects (including the side effects) of the drugs they administer.

In every sphere in which we take action and seek to achieve some result, the relation of cause and effect is fundamental. David Hume, one of the keenest of all thinkers in this arena, wrote:

All reasonings concerning matter of fact seem to be founded on the relation of *Cause and Effect*. By means of that relation alone we can go beyond the evidence of our memory and senses. If you were to ask a man, why he believes any matter of fact, which is absent; for instance, that his friend is in the country, or in France; he would give you a reason; and this reason would be some other fact; as a letter received from him, or the knowledge of his former resolutions and promises. A man finding a watch or any other machine in a desert island, would conclude that there had once been men in that island. All our reasonings concerning fact are of the same nature. . . . If we would satisfy ourselves, therefore, concerning the nature of that evidence, which assures us of matters of fact, we must enquire how we arrive at the knowledge of cause and effect."<sup>1</sup>

The methods by which we arrive at such knowledge are the central concern of this chapter. This matter is complicated, however, by the fact that there

are several different meanings of the word "cause." Therefore we begin by distinguishing these meanings from one another.

Things do not just happen. Events take place *under certain conditions*, and it is an axiom in the study of nature that to understand the world in which we live we must seek to learn the conditions under which events do or do not happen. It is customary to distinguish between the *necessary* and the *sufficient* conditions for the occurrence of an event.

A **necessary condition** for the occurrence of a specified event is a circumstance in whose *absence* the event *cannot* occur. For example, the presence of oxygen is a necessary condition for combustion to occur. If combustion occurs, then oxygen must have been present, because in the absence of oxygen there can be no combustion.

A **sufficient condition** for the occurrence of an event is a circumstance in whose *presence* the event *must* occur. The presence of oxygen is a necessary condition for combustion, as we noted, but it is not a sufficient condition for combustion to occur—because it is obvious that oxygen can be present without combustion occurring. For almost any substance, however, there is some range of temperature such that being in that range of temperature in the presence of oxygen is a sufficient condition for the combustion of that substance. So it is clear that for the occurrence of an event there may be several necessary conditions—and all of those necessary conditions must be included in the sufficient condition of that event.

Now the word "cause" is used (with respect to some event) sometimes to mean "the necessary condition of that event," and sometimes to mean "the sufficient condition of that event." It is most often used in the sense of necessary condition when the problem at hand is the *elimination* of some undesirable phenomenon. To eliminate it, one need only find some condition that is necessary to the existence of that phenomenon, and then eliminate that condition. What virus or bacterium is the cause of a certain illness? The physician cures the illness by administering a drug that will destroy those germs. The germs are said to be the *cause* of the disease in that they are a *necessary condition* for it—because in their absence the disease cannot occur.

However, the word "cause" is also commonly used to mean *sufficient condition*—especially when we are interested in the *production* of something desired, rather than the elimination of something undesirable. The metallurgist aims to discover what will produce greater strength in metal alloys, and when it is found that a certain process of mixed heating and cooling has that desired result, we say that such a process is the *cause* of the stronger alloy. It is correct to use the word "cause" in the one sense (necessary condition), or in the other (sufficient condition), but one should be clear about which of those meanings is intended.

Closely related to *sufficient condition* is another sense of the word "cause"—when a given phenomenon *tends* to have a causative role in the production of certain outcomes. For example, it is indeed correct to say that "smoking causes lung cancer," even though smoking cigarettes may long continue without having cancer as its result. And smoking is certainly not a necessary condition of lung cancer, because many such cancers arise in the total absence of smoking. But smoking cigarettes, in conjunction with very common biological circumstances, so frequently plays a role in the development of lung cancer that we think it correct to report that smoking is a "cause" of cancer.

This points to yet another common use of the term word "cause"—cause as the one factor that was critical in the occurrence of some phenomenon. An insurance company sends investigators to determine the cause of a mysterious fire. The investigators are likely to lose their jobs if they report that it was the presence of oxygen in the atmosphere that was the fire's cause—and yet of course it was (in the sense of necessary condition), for had there been no oxygen present there would have been no fire. Nor is the sufficient condition of the fire of interest to the company, for if the investigators reported that, although they had proof the fire was deliberately ignited by the policyholder, they had not yet been able to learn *all* the necessary conditions of the fire and therefore had not yet determined its full cause, they would certainly lose their jobs! What the company was seeking to discover was the incident or action that, in the presence of those conditions that usually prevail, *made the difference* between the occurrence and nonoccurrence of the fire.

In the real world, a huge man, forcibly resisting arrest, died shortly after having been beaten into submission by police officers in Cincinnati, Ohio, in November 2003. The county coroner investigating the death held it to be "homicide," carefully noting that hostile or malign intent is not implied by that word. "Absent the struggle," the coroner said, "Mr. Jones would not have died at that precise moment in time, and the struggle therefore is the primary cause of his death."<sup>2</sup> This sense of cause as "critical factor" is common and useful.

And there are subdivisions of this third sense of cause. When there is a causal sequence—a chain of events in which *A* causes *B*, *B* causes *C*, *C* causes *D*, and *D* causes *E*—we may regard the outcome, *E*, as the effect of any one of those preceding events. The death described above (symbolized by *E*) was caused by the struggle, the struggle (*D*) was caused by the resistance, the resistance (*C*) was caused by the arrest, the arrest (*B*) was caused by some violation of law (*A*), and so on. We distinguish between the **remote cause** and the **proximate cause** of *E*. The proximate cause is the event closest to it in the chain of events. The death, *E*, is the result of the proximate cause of the struggle, *D*; the others are remote: *A* more remote than *B*, *B* more remote than *C*, and so on.

Persons who leave school before the age of 16 are five times more likely to die from a heart attack as are university graduates; and the death rate within one year of a heart attack is 3.5 percent for college graduates but 20 percent for those with fewer than eight years of formal schooling.<sup>3</sup> But a college education is not the proximate cause of good health, nor is ignorance the proximate cause of disease. A poor education is a link in the causal chain, often resulting in a less adequate understanding of the disease process and thus a failure to make the lifestyle changes needed to promote better medical outcomes. So it is commonly and correctly observed that poverty, affecting education almost universally, is one of the "root causes" of poor health—not its proximate cause, of course—but a remote cause that needs uprooting.

The several different senses of the word "cause" need to be distinguished. We can legitimately infer cause from effect only when by *cause* we mean *necessary condition*. And we can infer effect from cause only when by *cause* we mean *sufficient condition*. When inferences are drawn both from cause to effect and from effect to cause, the word "cause" must be used in the sense of **necessary and sufficient condition**—the cause regarded as the sufficient condition of the event and that sufficient condition regarded as the conjunction of all its necessary conditions. No single definition of *cause* conforms to all the different (and reasonable) uses of that word.

## 12.2 Causal Laws and the Uniformity of Nature

---

Every use of the word "cause," whether in everyday life, or in science, involves or presupposes the doctrine that cause and effect are uniformly connected. We will allow that some particular circumstance was the cause of some particular effect only if we agree that any other circumstance of that type will (if the attendant circumstances are sufficiently similar) cause another effect of the same kind as the first. In other words, similar causes produce similar effects. As we use the word "cause," part of its meaning is that every occurrence of a cause producing some effect is an *instance* or *example* of the general causal law that such circumstances are always accompanied by such phenomena. If it can be shown that in another situation, after an occurrence of that supposed cause, the supposed effect did *not* occur, we will relinquish the belief that the one is the cause of the other.

Because every assertion that a particular circumstance was the cause of a particular phenomenon implies the existence of some causal law, every assertion of causal connection contains a critical element of *generality*. A **causal law**, as we use the term, asserts that a circumstance of such-and-such kind is invariably attended by a phenomenon of a specified kind, no matter where or when it occurs.



How can we come to know such general truths? The causal relation is not purely logical or deductive; as David Hume emphasized, it cannot be discovered by any *a priori* reasoning.\* Causal laws can be discovered only empirically, *a posteriori*, by an appeal to experience. But our experiences are always of *particular* circumstances, *particular* phenomena, and *particular* sequences of them. We may observe several instances of a circumstance (call it *C*), and every instance that we observe may be accompanied by an instance of a certain kind of phenomenon (call it *P*). But we will have experienced only some of the instances of *C* in the world, and our observations can therefore show us only that some cases of *C* are attended by *P*. Yet our aim is to establish a general, causal relation. How are we to get from the particulars we experience to the general proposition that *all* cases of *C* are attended by *P*—which is involved in saying that *C causes P*?

### 12.3 Induction by Simple Enumeration

When we assert that all cases of *C* are attended by *P*—that is, when we affirm a general causal relation—we have gone beyond analogy. The process of arriving at universal propositions from the particular facts of experience is called **inductive generalization**. Suppose we dip blue litmus paper into acid and it turns red. Suppose we do this three times, or ten times, always with the same result. What conclusion do we draw? By *analogy* we may draw a *particular* conclusion about what will happen to the color of the next piece of litmus paper we dip in acid—the fourth or the eleventh. Or we may draw a general conclusion about what will happen to *every* piece of blue litmus paper when it is dipped in acid. If we do the latter, it is an *inductive generalization* with which our argument concludes.

When the premises of an argument report a number of instances in which two attributes (or circumstances, or phenomena) occur together, we may infer

---

\*Hume wrote: "But to convince us that all the laws of nature, and all the operations of bodies without exception, are known only by experience, the following reflections may, perhaps, suffice. Were any object presented to us, and were we required to pronounce concerning the effect, which will result from it, without consulting past observation, after what manner, I beseech you, must the mind proceed in this operation? It must invent or imagine some event, which it ascribes to the object as its effect; and it is plain that this invention must be entirely arbitrary. The mind can never possibly find the effect in the supposed cause, by the most accurate scrutiny and examination. For the effect is totally different from the cause, and consequently can never be discovered in it. . . . A stone or piece of metal raised into the air, and left without any support, immediately falls; but to consider the matter *a priori*, is there anything we can discover in this situation which can beget the idea of a downward, rather than an upward, or any other motion, in the stone or metal? . . . In vain, therefore, should we pretend to determine any single event, or infer any cause or effect, without the assistance of observation and experience" (*An Enquiry Concerning Human Understanding*, 1748, sec. IV).

by analogy that some particular instance of one attribute will also exhibit the other attribute. By inductive generalization we might infer that every instance of the one attribute will also be an instance of the other. Inductive generalization of the form

Instance 1 of phenomenon *E* is accompanied by circumstance *C*.

Instance 2 of phenomenon *E* is accompanied by circumstance *C*.

Instance 3 of phenomenon *E* is accompanied by circumstance *C*.

Therefore every instance of phenomenon *E* is accompanied by circumstance *C*.

is an **induction by simple enumeration**. An induction by simple enumeration is very similar to an argument by analogy, differing only in having a more general conclusion.

Simple enumeration is often used in establishing causal connections. Where a number of instances of a phenomenon are invariably accompanied by a certain type of circumstance, it is only natural to infer the existence of a causal relationship between them. Since the circumstance of dipping blue litmus paper in acid is accompanied in all observed instances by the phenomenon of the paper turning red, we infer by simple enumeration that dipping blue litmus paper in acid is the cause of its turning red. The analogical character of such an argument is quite apparent.

Because of the great similarity between argument by simple enumeration and argument by analogy, similar criteria for appraisal apply to both. Some arguments by simple enumeration may establish their conclusions with a higher degree of probability than others. The greater the number of instances appealed to, the greater is the probability of the conclusion. The various instances or cases of phenomenon *E* accompanied by circumstance *C* are often called *confirming instances* of the causal law asserting that *C* causes *E*. The greater the number of confirming instances, the greater is the probability of the causal law—other things being equal. Thus the first criterion for analogical arguments also applies directly to arguments by simple enumeration.

In a historical report, simple enumeration can provide persuasive grounds for inferring a causal relationship. To illustrate, legislative acts designed to savage some individual or group temporarily out of favor, called bills of attainder, are known to endanger their advocates when the pendulum of political power swings. The accuser today becomes the victim tomorrow. Condemning such a bill of attainder (aimed at Thomas Osborne, Earl of Danbury) in the British House of Lords, the Earl of Carnarvon drove the point home in 1678 with the following enumeration:

My Lords, I understand . . . not a little of our English history, from which I have learnt the mischiefs of prosecutions such as these, and the ill fate of the prosecutors.

I shall go no further back than the latter end of Queen Elizabeth's reign, at which time the Earl of Essex was run down by Sir Walter Raleigh, and your Lordships well know what became of Sir Walter Raleigh. My Lord Bacon, he ran down Sir Walter Raleigh, and your Lordships know what became of my Lord Bacon. The Duke of Buckingham, he ran down my Lord Bacon, and your Lordships know what happened to the Duke of Buckingham. Sir Thomas Wentworth, afterwards Earl of Strafford, ran down the Duke of Buckingham, and you all know what became of him. Sir Harry Vane, he ran down the Earl of Strafford, and your Lordships know what became of Sir Harry Vane. Chancellor Hyde, he ran down Sir Harry Vane, and your Lordships know what became of the Chancellor. Sir Thomas Osborne, now Earl of Danby, ran down Chancellor Hyde.

What will now become of the Earl of Danby, your Lordships best can tell. But let me see that man that dare run the Earl of Danby down, and we shall soon see what will become of him.<sup>4</sup>

Rhetorically effective though this recounting of instances may be, it does not provide a trustworthy argument. The conclusion, that there is a causal connection between malicious accusation and subsequent destruction, appeals to six confirming instances—but the very nature of those instances prevents them from distinguishing between confirming instances of a genuine causal law and mere historical accidents.

The heart of the difficulty is this: The method of simple enumeration takes no account—*can* take no account—of exceptions to the causal law being suggested. Any alleged causal law may be overthrown by a single negative case, for any one disconfirming instance shows that what had been proposed as a “law” was not truly general. Exceptions *disprove* the rule—for an exception (or “negative instance”) is either one in which the alleged cause is found and is not followed by the alleged effect (in this historical case, a bill of attainder whose author did not suffer a like fate), or one in which the effect is encountered while the alleged cause is absent—where (using our earlier schema) *C* is present without *E*, or *E* is present without *C*. In an argument by simple enumeration there is no place for either of these; the only legitimate premises in such an argument are reports of instances in which *both* the alleged cause and the alleged effect are present.

Four hundred years ago Sir Francis Bacon, in *The Advancement of Learning* (1605), clearly identified the shortcomings of induction by simple enumeration. He wrote: “The induction that proceeds by simple enumeration is childish;\* its conclusions are precarious, and exposed to peril from a contradictory instance; and it generally reaches decision on too small a number of facts, and on those only that are on hand.”

It is thus a grave weakness of simple enumeration arguments that, if we confine ourselves to them exclusively, we will not look for, and are

therefore unlikely even to notice, the negative or disconfirming instances that might otherwise be found. For this reason, despite their fruitfulness and value in *suggesting* causal laws, inductions by simple enumeration are not at all suitable for *testing* causal laws. Yet such testing is essential; to accomplish it we must rely upon other types of inductive arguments—and to these we turn now.

## 12.4 Methods of Causal Analysis

---

The classic formulation of the methods central to all induction were given in the nineteenth century by John Stuart Mill (in *A System of Logic*, 1843). His systematic account of these methods has led logicians to refer to them as Mill's methods of inductive inference. The techniques themselves—five are commonly distinguished—were certainly not invented by him, nor should they be thought of as merely a product of nineteenth-century thought. On the contrary, these are universal tools of scientific investigation. The names Mill gave to them are still in use, as are Mill's precise formulations of what he called the "canons of induction." These techniques of investigation are permanently useful. Present-day accounts of discoveries in the biological, social, and physical sciences commonly report the methodology used as one or another variant (or combination) of these five techniques of inductive inference called **Mill's methods**: They are

1. The method of agreement
2. The method of difference
3. The joint method of agreement and difference

---

\*Induction by simple enumeration can mislead even the learned. The author of an account of the creation of the Oxford English Dictionary [*The Professor and the Madman*] reported his faithful belief in a curious causal claim: "Ever since I was 4 years old, I have said 'White Rabbits' at the very first moment of waking on every single first day of every single month that has passed. My mother, tucking me into bed one night, told me to do it, to bring good fortune; and since I have enjoyed fair good fortune for all of my subsequent days I have assumed that the acceptance of this moderate and harmless habit has had something to do with it, and so has reinforced my need to keep up the practice. . . . No fewer than 696 times I have maintained what has been a quite unvarying routine . . . when my eyes have snapped open with the first streaks of daylight, I have always uttered 'White Rabbits.'"

—Simon Winchester, "Oh Dear, Too Late for the White Rabbits,"  
*The New York Times*, 7 October 2006

4. The method of residues
5. The method of concomitant variation

We will examine each of these in turn, presenting Mill's classic statement of each (with one exception), followed by explication and illustration. These are the techniques on which science does and will rely in the search for causal laws.

## A. THE METHOD OF AGREEMENT

John Stuart Mill wrote:

If two or more instances of the phenomenon under investigation have only one circumstance in common, the circumstance in which alone all the instances agree, is the cause (or effect) of the given phenomenon.

This method goes beyond simple enumeration in that it seeks not only to discover the repeated conjunction of cause with effect, but also to identify the *only* circumstance, the *one* circumstance, that is invariably associated with the effect, or phenomenon, in which we are interested. This is an essential, and exceedingly common, tool of scientific inquiry. In searching for the cause of some deadly epidemic, for example, or in searching for the cause of some geological phenomenon, the epidemiologist or geologist will seek out the special circumstances that in every instance attend that result. In what way, they ask, do apparently differing sets of circumstances *agree*, where that result is produced?

Imagine, among the residents of some residence hall, a rash of stomach upsets, whose cause we must learn. The first line of inquiry naturally will be: What food or foods were eaten by *all* those who fell ill? Foods that were eaten by some but not all of those afflicted are not likely to be the cause of the outbreak; we want to know what circumstance can be found to be *common* to every case of the illness. Of course, what turns out to be common may not be a food; it may be the use of some infected utensil, or proximity to some noxious effluent, or other circumstance. But only when some circumstance is found in which *all* the cases of the illness agree are we on the way to the solution of the problem.

Schematically, the **method of agreement** may be represented as follows, where capital letters represent circumstances and lowercase letters denote phenomena:

A B C D occur together with w x y z.

A E F G occur together with w t u v.

Therefore A is the cause (or the effect) of w.

This method is particularly useful in identifying a *kind* of phenomenon, or a *range* of circumstances, whose investigation holds scientific promise. In



molecular genetics, for example, the search for the causes of some inherited disease can be greatly narrowed down using the method of agreement. Is there a common factor among families in which some specific disorder is prevalent? By examining the genetic makeup of such families, then closing in on those genetic factors that are found in such families but are not found commonly in others, the chromosome (and sometimes the site on that chromosome) where the inherited defect lies may be identified. This has proved a very effective method in tracing the cause of some diseases.

Similarly, the fluoridation of water in developed areas around the globe was the consequence of the discovery, more than half a century ago, that in cities where the rates of dental decay were unusually low, the one circumstance in common was an unusually high level of fluorine in the water supply. To confirm the causal connection, two cities of comparable size along the Hudson River—Newburgh and Kingston, New York—were closely studied in the 1940s; Newburgh's water was treated with fluoride, Kingston's had no fluoride. The statistics proved remarkable: Children in Newburgh showed a 70 percent reduction in cavities by the time they reached 14 years of age—and yet there were no differences between the two cities in rates of cancer, birth defects, or heart disease. The full explanation of this prevention of cavities could not be given at that time, but enough was known to justify the fluoridation of municipal water systems.

The method is widely powerful. A very promising development in the effort to help smokers break their addiction to nicotine was the discovery, reported in *Science* in 2007, that in a small number of persons who had suffered an injury to a particular region of their brains called the *insula*, the desire to smoke was immediately lost! Something in the *insula* appears to be a critical element in addiction. When statistical analyses of the data were completed, said a lead investigator from the University of Southern California, "it turned out that the likelihood of quitting smoking with ease after *insula* damage was 136 times higher than for damage anywhere else in the brain." A neuroscientist from the National Institute of Drug Abuse was enthusiastic: "To have any kind of variable produce this rate of quitting smoking is remarkable, to have it associated with a particular brain region is fantastic."<sup>5</sup> Thus, for addiction researchers, who are eager to apply the method of agreement to nicotine addiction, a major question has now become: "Can we learn to de-activate the *insula*?"<sup>6</sup>

In short, whenever we find a *single circumstance common to all instances* of a given phenomenon, we may rightly conclude that we have located at least the region of its cause.

The method of agreement has serious limitations, however. Looking chiefly to confirming instances, the method by itself is often insufficient to identify the cause being sought. The data available are seldom so conveniently

arranged as to permit the identification of one circumstance common to cases. And when inquiry reveals more than one circumstance common to cases, this technique alone cannot evaluate those alternative possibilities.

Although the presence of agreement between circumstance and phenomenon is often inconclusive, the *absence* of agreement may help us to determine what is *not* the cause of a phenomenon of interest. The method of agreement in essence eliminative; it points to the fact that circumstances arising in some of the cases, but not all of the cases, of the phenomenon in which we are interested, are not likely to be its cause. Those who argue against an alleged causal relation, therefore, are likely to call attention to the absence of uniform agreement, inferring that the alleged cause can be neither the sufficient condition nor the necessary condition of that phenomenon.

After we have learned all that the method of agreement can teach, other inductive methods capable of greater refinement in the search for causes are sure to be required.

### EXERCISES

Analyze each of the following scientific reports, explaining how the pattern of the method of agreement is manifested by each. Discuss, in each case, the limitations of the method of agreement as applied to that quest for a causal connection.

1. Contaminated scallions, chopped up raw in salsa that was served free to every table at a Chi-Chi's restaurant in western Pennsylvania, almost certainly caused the large outbreak of hepatitis A in the region, the Centers for Disease Control and Prevention said yesterday. Bunches of scallions (green onions) were stored together in large buckets for five days or more with the ice they had been shipped in from Mexico. As a result, even if only some bunches were tainted with the hepatitis virus when they were delivered, it would have quickly spread to all the other scallions—the ice water in the bucket becoming “hepatitis soup.” The scallions were later rinsed, chopped, refrigerated for two more days, and then added to the salsa which was made in 40-quart batches and kept refrigerated for up to three days. The outbreak, which has killed three people and made 575 other Chi-Chi's patrons sick, is the nation's biggest outbreak of hepatitis A from one source. Hepatitis A is spread by fecal matter from infected people, particularly those who fail to wash their hands after using the restroom. The virus does not multiply outside the body, but it can survive in food.

Hepatitis A is a common childhood disease in Mexico, and children commonly work on the scallion farms there; sewage-contaminated

water could also have been the culprit, whether used to irrigate the scallions, or wash them, or make the ice used in shipping. How the scallions became contaminated is not known.

—"Government Makes It Official: Blame Scallions for Outbreak,"  
*The New York Times*, 22 November 2003

2. Researchers at the University of California at Irvine have theorized that listening to Mozart's piano music significantly improves performance on intelligence tests. Dr. Frances H. Rauscher and her colleagues reported:

We performed an experiment in which students were each given three sets of standard IQ spatial reasoning tasks; each task was preceded by 10 minutes of

1. listening to Mozart's Sonata for Two Pianos in D major, K. 488; or
2. listening to a relaxation tape; or
3. silence.

Performance was improved for those tasks immediately following the first condition compared to the second two.

Test scores rose an average of 8 or 9 points following the Mozart sonata. Some of the students had reported that they liked Mozart, and some that they did not, but there were no measurable differences attributable to varying tastes. "We are testing a neurobiological model of brain function with these experiments," Dr. Rauscher said, "and we hypothesize that these patterns may be common in certain activities—chess, mathematics, and certain kinds of music. . . . Listening to such music may stimulate neural pathways important to cognition."

—Frances H. Rauscher, Gordon L. Shaw, Katherine N. Ky,  
"Music and Spatial Task Performance," *Nature*, 14 October 1993

3. Medical researchers have concluded not only that the timing of sexual intercourse in relation to ovulation strongly influences the chance of conception, but that conception occurs *only* when intercourse takes place during a specifiable period in the menstrual cycle. The researchers summarized their findings thus:

We recruited 221 healthy women who were planning to become pregnant. At the same time the women stopped using birth control methods, they began collecting daily urine specimens and keeping daily records of whether they had sexual intercourse. We measured estrogen and progesterone metabolites in urine to estimate the day of ovulation.

In a total of 625 menstrual cycles for which the dates of ovulation could be estimated, 192 pregnancies were initiated. . . . Two thirds ( $n = 129$ ) ended in live births. Conception occurred only when intercourse took place during a six-day period that ended on the estimated day of ovulation. The probability of conception ranged from 0.10 when

intercourse occurred five days before ovulation to 0.33 when it occurred on the day of ovulation itself.

Conclusion: Among healthy women trying to conceive, nearly all pregnancies can be attributed to intercourse during a six-day period ending on the day of ovulation.

—Allen J. Wilcox, Clarice R. Weinberg, Donna D. Baird,  
 “Timing of Sexual Intercourse in Relation to Ovulation,”  
*The New England Journal of Medicine*, 7 December 1995

4. A large extended family in the town of Cartago, Costa Rica, has long suffered an unusual affliction—an incurable form of genetically caused deafness. Children born into the family have a 50 percent chance of developing the disease, and learn their fate at about the age of ten, when those who have inherited a genetic mutation find that they are beginning to lose their hearing. Scientists from the University of Washington have recently traced the cause of the family’s affliction to a previously unknown gene, named the diaphanous gene, that helps operate the delicate hair cells in the inner ear that respond to sound vibrations.

This gene has a single mutation appearing in the Costa Rican family, whose founder arrived in Cartago from Spain in 1713, and who suffered from this form of deafness—as have half his descendants in the eight generations since. Many in the family remain in Cartago because the family’s hereditary deafness is well-known and accepted there. With only a single family to be studied, and thus very few genetic differences to work with, pinpointing the gene took six years. The critical mutation involved just one of the 3,800 chemical letters that constitute the gene’s DNA.

—Reported in *Science*, 14 November 1997

5. Researchers from the National Cancer Institute announced that they have found a number of genetic markers shared by gay brothers, indicating that homosexuality has genetic roots. The investigators, reporting in *Science*, 16 July 1993, have found that out of 40 pairs of gay brothers examined in their study, 33 pairs shared certain DNA sequences on their X chromosome, the chromosome men inherit only from their mothers. The implicit reasoning of this report is that, if brothers who have specific DNA sequences in common are both gay, these sequences can be considered genetic markers for homosexuality.
6. The relation between male circumcision and HIV infection has been a concern of the British medical journal, *Lancet*, for many years. Before the turn of this century investigators studying that relation wrote, in

*Lancet*, that studies going back as far as 1989 showed a very greatly increased risk of HIV-1 infection for men who are not circumcised. The epidemiological and biological evidence that links the two, they later wrote, "has become compelling." Very recent studies in Kenya and Uganda have produced evidence that is even more compelling. In 2006, trials in those countries conducted by the U.S. National Institutes of Health were *stopped* because the results were so clear! It appeared that circumcision reduces a man's risk of contracting AIDS from heterosexual sex by about half, and therefore U.S. officials concluded that it would have been unethical to continue without offering circumcision to all 8000 men in the trials. The final figures, reevaluated and published in *Lancet* on 23 February 2007, are even more striking. They suggest that *circumcision reduces a man's risk of contracting AIDS by as much as 65 percent*. Dr. Anthony Fauci, of the National Institute of Allergy and Infectious Diseases, was emphatic: "Look. This is a one-time, permanent intervention that's safe when done under appropriate medical conditions. If we had an AIDS vaccine that was performing as well as this, it would be the talk of the town."

## B. THE METHOD OF DIFFERENCE

John Stuart Mill wrote:

If an instance in which the phenomenon under investigation occurs and an instance in which it does not occur, have every circumstance in common save one, that one occurring only in the former, the circumstance in which alone the two instances differ, is the effect, or the cause, or an indispensable part of the cause, of the phenomenon.

This pattern focuses not on what is common among those cases in which the effect is produced, but on what is *different* between those cases in which the effect is produced and those in which it is not. If we had learned, when investigating that rash of stomach upsets described earlier, that all those who had become ill had eaten the canned pears for dessert, but that the pears had been eaten by none of those who did not become ill, we would be fairly confident that the cause of the illness had been identified.

The difference between the *method of difference* and the method of agreement is highlighted in a recent report about the role of the hormone testosterone in the aggressive conduct of males.

Among many species, testes are mothballed most of the year, kicking into action and pouring out testosterone only during a very circumscribed mating season—precisely the time when male–male aggression soars. Impressive though they seem, these data are only correlative—[reporting only] testosterone found on the scene repeatedly when aggression has occurred.



The proof comes with the knife, the performance of what is euphemistically known as a subtraction experiment. Remove the source of the testosterone in species after species, and levels of aggression plummet. Reinstall normal testosterone levels afterward with injections of synthetic testosterone and aggression returns.

The subtraction and replacement paradigm gives damning proof that this hormone is involved in aggression.<sup>7</sup>

Testosterone makes the critical difference, clearly, but the author of this report is careful not to assert that testosterone is *the cause* of male aggression. More accurately, the report states that testosterone is surely *involved* in aggression. As Mill would put it, the hormone is *an indispensable part of the cause* of male aggression. Wherever we can identify a single factor that makes the critical difference when all else remains normal—the factor that eliminates the phenomenon in question when we remove it, or the factor that produces the phenomenon in question when we introduce it—we will pretty surely have identified the cause, or an indispensable part of the cause, of the phenomenon we are investigating.

Schematically, where again capital letters denote circumstances and lowercase letters denote phenomena, the **method of difference** may be represented as follows:

A B C D occur together with w x y z.

B C D occur together with x y z.

Therefore A is the cause, or the effect, or an indispensable part of the cause of w.

The method of difference is of central importance in scientific investigations of almost every kind. One vivid illustration of its use is the ongoing investigation by medical researchers into the effects of particular proteins suspected as being implicated in the development of certain diseases. Whether the substance under investigation really is the cause (or an indispensable part of the cause) can only be determined when we create an experimental environment in which that substance has been eliminated. Investigators sometimes are able to do just that—not in humans, of course, but in mice which are subject to the same disease, from whom the gene that is known to produce the suspect protein is deleted. Animals so treated are then inbred, creating populations of what are called “knockout mice,” precious in the world of contemporary medical research, in which the process relevant to the disease in question can be studied in an animal exactly like other animals subject to that disease, *except for the critical difference created by the knockout*, the absence of the substance hypothesized as cause. Such studies have resulted in some remarkable medical advances.

To illustrate, using knockout mice, scientists have been able to identify the gene that causes inflammation—swelling, redness, and pain. The gene *MIP-1 alpha*, present in mice and in humans, was suspected of producing the protein

that begins the process of inflammation. Pathologists at the University of North Carolina at Chapel Hill bred mice that *lacked* the gene *MIP-1 alpha*, and then infected those mice, and also a control group of normal mice, with viruses known to cause influenza and other diseases. The normal mice did develop extreme inflammation as expected, but mice lacking the *MIP-1 alpha* gene had only slight inflammation. This is one big step toward the development of drugs that will allow humans to fight viral infections without painful and damaging inflammation.<sup>8</sup>

A famous and very dramatic illustration of the method of difference is provided by the following account of experiments confirming the true cause of yellow fever, long one of the great plagues of human kind. The experiments described here were conducted by U.S. Army doctors Walter Reed, James Carroll, and Jesse W. Lazear in November 1900. Earlier that year Dr. Carroll had contracted yellow fever by deliberately allowing himself to be bitten by an infected mosquito in another experiment; soon after, Dr. Lazear died of yellow fever, and the camp in which the following experiments took place was named for him:

Experiments were devised to show that yellow fever was transmitted by the mosquito alone, all other reasonable opportunities for being infected being excluded. A small building was erected, all windows and doors and every other possible opening being absolutely mosquito-proof. A wire mosquito screen divided the room into two spaces. In one of these spaces fifteen mosquitoes, which had fed on yellow fever patients, were liberated. A nonimmune volunteer entered the room with the mosquitoes and was bitten by seven mosquitoes. Four days later, he suffered an attack of yellow fever. Two other nonimmune men slept for thirteen nights in the mosquito-free room without disturbances of any sort.

To show that the disease was transmitted by the mosquito and not through the excreta of yellow fever patients or anything which had come in contact with them, another house was constructed and made mosquito-proof. For 20 days, this house was occupied by three nonimmunes, after the clothing, bedding and eating utensils and other vessels soiled with the discharge, blood and vomitus of yellow fever patients had been placed in it. The bed clothing which they used had been brought from the beds of the patients who had died of yellow fever, without being subjected to washing or any other treatment to remove anything with which it might have been soiled. The experiment was twice repeated by other nonimmune volunteers. During the entire period all the men who occupied the house were strictly quarantined and protected from mosquitoes. None of those exposed to these experiments contracted yellow fever. That they were not immune was subsequently shown, since four of them became infected either by mosquito bites or the injection of blood from yellow fever patients.<sup>9</sup>

That portion of the experiment described in the first paragraph above very deliberately created a single important difference between the subjects in the two carefully enclosed spaces: the presence of mosquitoes that had fed on yellow fever patients in the one space, the absence of such mosquitoes in the other.

That portion of the experiment described in the second paragraph above deliberately created a second use of the method of difference, in which the only significant difference between two groups of subjects, both of whom had submitted to very close contact with yellow fever victims, was the exposure of some of them to infected mosquito bites or infected blood. Absent that circumstance, no infection arose.

Science seeks causal laws. In the never-ending efforts to confirm or to disconfirm hypothesized causal connections, the method of difference is pervasive and powerful.

## EXERCISES

Analyze each of the following reports, explaining the ways in which the method of difference has been applied in the investigations recounted. Discuss the strengths and weaknesses of the method of difference as it is used in each case.

1. How critical is sleep to memory? Researchers at two universities, separately, conducted experiments in 2003 designed to determine how sleep affects our ability to remember. College-age people were trained to perform certain tasks and then tested to see how much they recalled in confronting such tasks after either a night's sleep or several hours awake. "We all have the experience of going to sleep with a question and waking up with the solution," observed one of the investigators, Prof. Danial Margoliash, of the University of Chicago. But does the sleep really help?

It does, markedly. Not just as a matter of re-charge, but, the investigators found, because sleep *rescues* memories by storing and consolidating them deep in the brain's circuitry. At the University of Chicago, subjects trained to understand murky speech on a voice synthesizer could regularly understand more words after a night of sleep than matched counterparts who were tested just hours after the training with no intervening sleep. And at the Harvard Medical School, one hundred subjects were trained to perform certain finger-tapping sequences that they were later asked, at various intervals, to repeat. The process of memory consolidation required one or two nights of sleep—after which the performance of the subjects improved substantially.

— Reported in *Nature*, 9 October 2003

2. The heavy use of salt is widely suspected by experts to be the cause of an epidemic of high blood pressure and many deaths from heart

disease around the world. But how prove that salt is the culprit? There are "natural experiments" in which isolated jungle or farming communities are introduced to modern civilization, move to cities, adopt high-salt diets, and commonly develop high blood pressure. But such evidence is inconclusive because many important factors change together; new stresses and many dietary changes accompany the increase in salt. How can the causal effects of salt by itself be tested?

Dr. Derek Denton, of the University of Melbourne, selected a group of normal chimpanzees, a species biologically very close to humans, in which to conduct the needed trials. A group of chimpanzees in Gabon, with normal blood pressure, were first studied in their natural state. The group was then divided in half, with one half receiving gradually increasing amounts of salt in their diet for twenty months. Normal blood pressure in a chimpanzee is 110/70. In Dr. Denton's experiment, the animals' blood pressure commonly rose as high as 150/90, and in some individuals much higher. But among animals in the control group, who received no additional salt, blood pressure did not rise. Six months after the extra salt was withdrawn from their diet, all the chimpanzees in the experimental group had the same low blood pressure they had enjoyed before the experiment. Because there was no other change in the lifestyle of those animals, the investigators concluded that changes in salt consumption caused the changes in blood pressure.

—D. Denton et al., "The Effect of Increased Salt Intake on Blood Pressure of Chimpanzees," *Nature*, October 1995

3. Does Louisiana hot sauce, the principal ingredient of the spicy New Orleans cocktail sauce commonly served with raw shellfish, kill certain bacteria found in raw oysters and clams? The answer appears to be yes. Bacteria of an infectious and sometimes fatal kind—*Vibrio vulnificus*—are found in 5 to 10 percent of raw shellfish on the market. Dr. Charles V. Sanders and his research team, from Louisiana State University Medical Center in New Orleans, added Louisiana hot sauce to cultures of *Vibrio* growing in test tubes; the sauce, even when greatly diluted, killed *V. vulnificus* in five minutes or less. "I couldn't believe what happened," Dr. Sanders said, admitting that he still eats raw oysters, "but only with plenty of hot sauce."

—Reported to the Interscience Conference on Antimicrobial Agents, New Orleans, October 1993

4. In Lithuania, rear-end auto collisions happen as they do in the rest of the world; bumpers crumple, tempers flare. But drivers there do

not seem to suffer the complaints so common in the United States, the headaches and lingering neck pains known as “whiplash syndrome.” Dr. Harald Schrader and colleagues from University Hospital in Trondheim, Norway, without disclosing the purpose of their study, gave health questionnaires to 202 Lithuanian drivers whose cars had been struck from behind one to three years earlier in accidents of varying severity. The drivers’ reports of their symptoms were compared to the reports of a control group (of the same size, same ages, and same home towns) of drivers who had not been in an accident. Thirty-five percent of the accident victims reported neck pain, but so did 33 percent of the controls; 53 percent of those who had been in an accident had headaches, but so did 50 percent of those in the control group. The researchers concluded: “No one in the study group had disabling or persistent symptoms as a result of the car accident.”

What, then, can account for the explosion of whiplash cases elsewhere in the world? Drivers in the Lithuanian study did not carry personal injury insurance at the time of the study, and people there very infrequently sue one another. Most medical bills are paid by the government, and at the time of the study there were no claims to be filed, no money to be won, and nothing to be gained from a diagnosis of chronic whiplash. Chronic whiplash syndrome, the Norwegian researchers concluded, “has little validity.”

—Harald Schrader et al., “Natural Evolution of Late Whiplash Syndrome Outside the Medicolegal Context,” *The Lancet*, 4 May 1996

5. To determine the role of specific genes, mice are bred in which certain genes have been deleted, called “knockout mice.” When normal mice are placed in a lighted room, with dark corners, they go immediately to the dark. In one recent experiment the mice, upon entering the dark, encounter a mild electric shock, and very quickly learn to stay away from those dark regions. Mice who lack a gene called *Ras-GRF* learn to be wary just as quickly as do normal mice. But, unlike normal mice, the knockout mice throw caution to the winds the next day, and chance the dark corners again and yet again. It appears that the *Ras-GRF* gene—probably very much like the analogous gene in humans—plays a critical role in the ability of the mice to remember fear. This gene is almost certainly crucial for the survival of mammals.

—Reported in *Nature*, December 1997

6. Here is some reassuring news for those whose career plans are slightly behind schedule: It turns out that peaking too early may kill you.



That's the finding of Stuart J. H. McCann, a professor of psychology at the University College of Cape Breton in Nova Scotia.

McCann's research concerns what he calls the "precocity-longevity hypothesis." McCann analyzed the lives of 1,672 U.S. governors who served between 1789 and 1978 and found that those who were elected at relatively tender ages generally died earlier than their less precocious counterparts. Even when he controlled for the year that the governors were born, or how long they served, and what state they governed, the pattern held. No matter how he sliced the data, or ran the regressions, or accounted for various statistical biases, the story remained the same: governors elected to office at younger ages tended to have shorter lives.

And what holds for state executives seems also to hold for other young achievers. McCann also analyzed smaller but more diverse sets of accomplished people—including American and French presidents, Canadian and British prime ministers, Nobel Laureates, signers of the Declaration of Independence, Academy Award winners, and seven centuries of Roman Catholic pontiffs. Again he found that "those who climb to the loftiest peaks in the shortest time also die younger. For the eminent, and perhaps for all, an early rise may lead to an early fall."

—*Personality and Social Psychology Bulletin*, February 2003

7. Cholera, caused by a water-borne bacterium ingested by drinking contaminated water, is a dreadful disease; pandemics of cholera in the 19th century killed tens of thousands. The accepted view, that it was caused by breathing a filthy miasma, was doubted by John Snow, a founding member of the London Epidemiological Society. When a terrible cholera epidemic struck London in 1848-49, Snow hypothesized that bad water, from urban wells and from the Thames River, was the villain. Some water companies drew their water from within the tidal section of the Thames, where the city's sewage was also dumped, thus providing their customers with excrement-contaminated drinking water. It stank, so some of the intake pipes were shifted to points above the tideway. In 1854 cholera returned with even greater horror. Snow identified two water companies, one of which had moved its intake to a point above the tidal region of the river, the other still supplying a fecal cocktail; his data from these two districts showed a strong connection between cholera mortality and water source. Snow also identified a particular well, on Broad Street, and plotted cholera mortality house by house in the area of that well—the number of dead increasing sharply with proximity to the Broad Street pump—while a

few streets away on Warwick Street there were no cholera deaths at all. Just across from the Broad Street pump was the Poland Street Workhouse, whose wretched inmates remained healthy—the workhouse had its own well. The workers of the Lion Brewery, close to the pump on Broad Street, also had its own well; its workers did not contract cholera—they drank mainly malt liquor. The outbreak ended when Snow persuaded the authorities to *remove the handle from the Broad Street pump*. There is today a replica of the handleless pump outside a nearby pub named in honor of John Snow.

Steven Shapin, "Sick City," *The New Yorker*, 6 November 2006

### C. THE JOINT METHOD OF AGREEMENT AND DIFFERENCE

Although Mill believed that the **joint method of agreement and difference** was an additional and separate technique, it is best understood as the combined use of the method of agreement and the method of difference in the same investigation. It can be represented schematically (capital letters again denoting circumstances, lowercase letters denoting phenomena) as follows:

$$\begin{array}{ll} ABC - xyz. & ABC - xyz. \\ ADE - xtw. & BC - yz. \end{array}$$

---

Therefore A is the effect, or the cause, or an indispensable part of the cause, of x.

Because each of the two methods (agreement schematized above on the left, difference schematized on the right) affords some probability to the conclusion, their joint use affords a higher probability to that conclusion. In many scientific investigations this combination serves as an extremely powerful pattern of inductive inference.

A notable advance in medicine provides an illustration of the power of the joint method. Hepatitis A is a liver infection that afflicted tens of thousands of Americans; it spread widely among children, chiefly through contaminated food or water, and was sometimes deadly. How might it be prevented? The ideal solution, of course, would be an effective vaccine. However, an enormous difficulty faced those who would test any vaccine for hepatitis A: It was very hard to predict where outbreaks of the infection would occur, and therefore it was usually not possible to select experimental subjects in ways that would yield reliable results. This difficulty was finally overcome in the following way.

A potential vaccine was tested in a community of Hasidic Jews, Kiryas Joel, in Orange County, New York, a community that was highly unusual in that it was plagued by yearly epidemics of this infection. Almost no one escaped hepatitis A in Kiryas Joel, and nearly 70 percent of the community members had been infected by the time they are nineteen years old. Dr. Alan

Werzberger, of the Kiryas Joel Institute of Medicine, and his colleagues recruited 1037 children in that community, ages two to sixteen, who had not been exposed to the hepatitis A virus, as determined by a lack of antibodies to the virus in their blood. Half of them (519) received a single dose of the new vaccine, and among those vaccinated children not a single case of hepatitis A was reported. Of the 518 children who received dummy injections, 25 became infected with hepatitis A soon after. The vaccine for hepatitis A had been found.<sup>10</sup>

Liver specialists in Boston and Washington greeted this study with admiration, calling it "a great breakthrough" and a "major medical advance." What is the pattern of inference on which this achievement relied? Both the method of agreement and the method of difference were employed, as is common in medical investigations. Among all those young residents of the community who became immune to hepatitis A, there was only one relevant circumstance *in common*: All the immunes had received the new vaccine. By itself, this strongly tended to show that the vaccine did cause that immunity. The method of difference supported this conclusion overwhelmingly: The circumstances of those who did become immune and those who did not were essentially alike *in every respect except one*, the administration of the vaccine to the immune residents.

The testing of new drugs or procedures is often conducted in what are called "double-arm" trials, one group receiving the new treatment while the other group does not, after which (in suitable cases) there may be a carefully executed crossover, in a second phase, in which those who originally did not receive the treatment do so, and those who originally did receive the treatment do not. The application of the joint method of agreement and difference underlies such investigations, which are common and exceedingly productive.

## EXERCISES

Analyze each of the following reports, explaining the way in which the method of agreement and the method of difference have been jointly applied, and identifying the special force, if any, of their combination.

1. Pain can be agonizing, but it serves a useful function: It teaches people and animals to avoid dangers, and forces them to attend to wounds. Strangely, there are a very few people who never feel pain; they remain unaware of having suffered significant injuries.

One family in northern Pakistan has several such members. One, a ten-year old boy, became famous for giving street performances in which he put knives through his arms and walked on hot coals.

Tissue damage would result, but no discomfort. Geneticist C. G. Woods of Cambridge University searched for the cause of this remarkable inability to feel pain. Eventually he zeroed in on mutations in a gene, *SCN9A*, which codes for the channel through which sodium enters pain-sensing cells, critical to the pain signal. Testing with electric current, he could open and close sodium channels on some cells—but he could not open the sodium channels on those mutant cells. Said Woods, “This shows that rare diseases can still be of great importance because of the insights they give into biological processes.”

A Yale University neurologist, Stephen Waxman, observes that if researchers could craft a drug that can make these channels inactive, as they are in the Pakistani family members, millions of people worldwide who suffer from chronic pain would be wonderfully served.

—Reported in *Nature*, 14 December 2006

2. A deadly heart ailment affecting about 1 million African American men—familial amyloid cardiomyopathy—and another that afflicts older men of all ethnicities, are known to be caused by an abnormally folded protein that builds up in the organism. Transthyretin protein, made in the liver, has four subunits. A mutation in the gene that makes two of those subunits results in the instability of the protein, its misfolding, and eventually in death. That this is indeed the cause of the ailments was shown by the fact that a liver transplant, providing a healthy version of the critical gene, can result in cure—but often that correction comes too late to stymie the misfolding that did the damage.

A strange twist of nature, reported in *Science* in January 2003 by Dr. Jeffrey Kelly, of the Scripps Research Institute in San Diego, provided the clues to a therapy that can thwart the misfolding process. Because diseases of this kind are quite common in Portugal, families there are screened to see who has the mutated gene and is therefore at risk. One very large family was identified whose members had the mutated gene and yet never did contract the disease. It turned out that in this family, a second gene that made the other two subunits of the protein had undergone its own mutation, suppressing or reversing the disease process. Members of that family carried a cure to an inherited disease in their own genes.

Dr. Kelly found that as a result of this further mutation the disease was prevented by the erection of a kind of barrier between the normal and the abnormal protein states. Then, by screening libraries of small molecules, he located several that, already approved by the Food and

Drug Administration for other purposes, could mimic the effect of the second mutation, successfully reversing the misfolded protein in animals.

3. Sixteen-year-old David Merrill, of Suffolk, Virginia, hypothesized that the loud sounds of hard-rock music have a bad effect on its devoted fans. He tested the theory on mice. Seventy-two mice were divided into three groups of 24, the first to be exposed to hard-rock music, the second to music by Mozart, and the third to no music at all. After becoming accustomed to their environments, but before being exposed to the music, Merrill tested all the mice in a maze, which took them an average of 10 minutes to complete. Then the groups were exposed to the music for 10 hours a day.

With repeated testing the control-group mice *reduced* their time in the maze by an average of 5 minutes. Those exposed to Mozart *reduced* their time by 8.5 minutes. The hard-rock mice *increased* their time in the maze by 20 minutes.

Merrill also reported that when, in an earlier attempt, he had allowed all the mice to live together, the project had to be cut short because, unlike the Mozart-listening mice, the hard-rock-listening mice killed other mice.

—Reported in *Insight*, 8 September 1997

4. Scientists have long known that severely restricting the number of calories that mice and other organisms consume lengthens their life span. Animals on low-calorie diets typically have abnormally cool body temperatures. Does low temperature, in itself, result in longer life? The answer is yes.

Bruno Conti, of the Scripps Research Institute in La Jolla, California, genetically engineered mice to have a faulty sense of body temperature. The alteration reduced the animals' temperatures by 0.03 to 0.05°C below normal; they were given as much food as they wanted, maintaining their normal weight. The low-temperature mice lived about 15 percent longer than normal mice did.

—Reported in *Science*, 3 November 2006

5. At a social gathering of eighty-five faculty members, graduate students, and staff workers in the Department of Food Science at the University of Illinois in Urbana-Champaign, the partygoers served themselves ice cream. They did not know they were also the subjects of an experiment. Half the participants were given 17-ounce bowls, and half 34-ounce bowls. In addition, half were given 2-ounce spoons to scoop out their ice cream, and half were given 3-ounce serving spoons.



With larger spoons, people served themselves 14.5 percent more, and with a larger bowl they heaped on 31 percent more. With both large spoon and large bowl these nutrition experts helped themselves to 56.8 percent more ice cream than those who used the smaller utensils. And all but three ate every bit of the ice cream they took. Smaller platters and smaller utensils may be the key to a successful diet.

—Reported by Brian Wansink in *The American Journal of Preventive Medicine*, September 2006

## D. THE METHOD OF RESIDUES

John Stuart Mill wrote:

Subduct from any phenomenon such part as is known by previous inductions to be the effect of certain antecedents, and the residue of the phenomenon is the effect of the remaining antecedents.

The first three methods seem to suppose that we can eliminate or produce the cause (or effect) of some phenomenon in its entirety, as indeed we sometimes can. In many contexts, however, we can only deduce the causal effect of some phenomenon by observing the *change* that it makes in a set of circumstances whose cause is already understood in part.

This method, focusing on *residues*, is well illustrated by the very simple device used to weigh truck cargos. The weight of the truck when empty is known. To determine the weight of the cargo, the entire truck is weighed with its cargo—and the weight of the cargo is then known to be the weight of the whole minus the weight of the truck. The known “antecedent,” in Mill’s phrase, is the recorded weight of the empty truck that must be subtracted from the reading on the scale; the cause of the difference between that reading and the known antecedent is obviously attributable to the remaining “antecedents”—that is, to the cargo itself.

Schematically, the **method of residues** can be represented as follows:

$A \ B \ C \rightarrow x \ y \ z.$

$B$  is known to be the cause of  $y$ .

$C$  is known to be the cause of  $z$ .

Therefore  $A$  is the cause of  $x$ .

A splendid illustration of the effectiveness of the method of residues is provided by one of the great chapters in the history of astronomy, the discovery of the planet Neptune:

In 1821, Bouvard of Paris published tables of the motions of a number of planets, including Uranus. In preparing the latter he had found great difficulty in making an orbit calculated on the basis of positions obtained in the years after 1800

agree with one calculated from observations taken in the years immediately following discovery. He finally disregarded the older observations entirely and based his tables on the newer observations. In a few years, however, the positions calculated from the tables disagreed with the observed positions of the planet and by 1844 the discrepancy amounted to 2 minutes of arc. Since all the other known planets agreed in their motions with those calculated for them, the discrepancy in the case of Uranus aroused much discussion.

In 1845, Leverrier, then a young man, attacked the problem. He checked Bouvard's calculations and found them essentially correct. Thereupon he felt that the only satisfactory explanation of the trouble lay in the presence of a planet somewhere beyond Uranus which was disturbing its motion. By the middle of 1846 he had finished his calculations. In September he wrote to Galle at Berlin and requested the latter to look for a new planet in a certain region of the sky for which some new star charts had just been prepared in Germany but of which Leverrier apparently had not as yet obtained copies. On the twenty-third of September Galle started the search and in less than an hour he found an object which was not on the chart. By the next night it had moved appreciably and the new planet, subsequently named Neptune, was discovered within  $1^\circ$  of the predicted place. This discovery ranks among the greatest achievements of mathematical astronomy.<sup>11</sup>

The phenomenon under investigation here is the movement of Uranus. A great part of that phenomenon, the orbit of Uranus around the sun, was well understood at the time. Observations of Uranus approximated this calculated orbit but exhibited a puzzling residue, some perturbation of what had been calculated, for which further explanation was needed. An additional "antecedent"—that is, an additional existing factor that would account for the perturbation—was hypothesized to be another (undiscovered) planet whose gravity would, together with what was already known about the orbit of Uranus, explain that residue. Once hypothesized, that new planet, Neptune, was very quickly found.

The method of residues differs from the other methods in that it can be used with the examination of only one case, whereas the others require the examination of at least two cases. And the method of residues, unlike the others, appears to depend on antecedently established causal laws, while the other methods (as Mill formulated them) do not. The method of residues is nevertheless an inductive, not a deductive, method (as some have suggested), because it yields conclusions that are only probable and cannot be *validly deduced* from their premises. An additional premise or two might transform an inference by the method of residues into a valid deductive argument, but that can be said for other inductive methods as well.

## EXERCISES

Analyze each of the following arguments in terms of "antecedents" and "phenomena" to show how they follow the pattern of the method of residues.

1. For nineteen years space scientists, astronomers, and physicists have been puzzled by what appears to be a mysterious force pulling spacecraft in the direction of the sun. It was first noticed when the trajectories of two outward bound and very distant spacecraft (Pioneer 10 and 11, launched in 1972 and 1973) were carefully analyzed. The trajectories of two later probes (Galileo, launched toward Jupiter in 1989, and Ulysses, launched into polar orbit around the sun) have exhibited the same peculiarities: They give evidence of a weak force that perturbs their directions and velocities. This force was discovered by adding up the effects of all other known forces acting on the spacecraft and finding that something unexplained was left over.

This force is apparently slowing the outward progress of the spacecraft speeding away from or around the sun—but in contrast to the force of gravity, the strength of this mystery force does not decline proportionally to the inverse square of a spacecraft's distance from the sun, but instead at a linear rate, which makes it very unlikely that the mystery force is a gravitational effect of the sun.

Calculations were made using two independent methods, and data of different types, taking into account possible errors in the software and the hardware used in the measurements. A host of other possible errors were investigated and accounted for—and after ruling all of these out, a team of physicists from the Los Alamos National Laboratory announced that the mystery remained. This means that some hitherto unknown phenomenon may be at work—what physicists excitedly call “new physics.”

—Reported in *Physical Review Letters*, September 1998

2. In H. Davies' experiments on the decomposition of water by galvanism, it was found that besides the two components of water, oxygen and hydrogen, an acid and an alkali were developed at opposite poles of the machine. Since the theory of the analysis of water did not give reason to expect these products, their presence constituted a problem. Some chemists thought that electricity had the power of producing these substances of itself. Davies conjectured that there might be some hidden cause for this part of the effect—the glass might suffer decomposition, or some foreign matter might be in the water. He then proceeded to investigate whether or not the diminution or total elimination of possible causes would change or eliminate the effect in question. Substituting gold vessels for glass ones, he found no change in the effect and concluded that glass was not the cause. Using distilled water, he found a decrease in the quantity of acid and alkali involved, yet

enough remained to show that the cause was still in operation. He inferred that impurity of the water was not the sole cause, but was a concurrent cause. He then suspected that perspiration from the hands might be the cause, as it would contain salt which would decompose into acid and alkali under electricity. By avoiding such contact, he reduced the quantity of the effect still further, till only slight traces remained. These might be due to some impurity of the atmosphere decomposed by the electricity. An experiment determined this to be the case. The machine was put under an exhaust receiver and when it was thus secured from atmospheric influences, no acid or alkali was produced.

—G. Gore, *The Art of Scientific Discovery*, 1878

3. Satellite observations collected between 1992 and 2001 suggest that the upper surface of the Larsen C ice shelf, in Antarctica, dropped as much as 27 cm per year during that period. About a quarter of that shrinkage, or 7 cm, may have resulted from snow packing down into denser material called firn. Uncertainties about such factors as the height of the ocean tides, and the salinity of water beneath the ice shelf would account for no more than a small fraction of the remaining loss of height above water.

Therefore, concluded Andrew Shepherd, a glaciologist at the University of Cambridge in England, as much as 20 cm per year of the upper surface's drop must stem from melting. Nine-tenths of any mass of floating ice lies below the water's surface, suggesting that the Larsen C ice shelf is thinning by as much as 2 m each year.

The likely cause of this thinning is relatively warm water beneath the shelf. Even a very small temperature increase in the water below an ice shelf can make a big difference in the melting rate of the overlying ice. Larsen C is stable, and isn't shedding more icebergs than normal, Shepherd reported, but at its current rate of thinning, Larsen C could reach 200 m in thickness (the thickness at which other ice shelves have disintegrated) and therefore be susceptible to disintegration in 70 years—but if the waters in the region continue to warm, the demise of Larsen C could occur even sooner.

—Reported in *Science News*, 1 November 2003

4. Analyzing more than forty years of weather data, climatologists at the National Oceanic and Atmospheric Administration in Boulder, Colorado, recently found that the daily temperature range—the difference between the daytime maximum and the nighttime minimum temperatures—at 660 weather stations in the continental United States fluctuates in a very puzzling manner: the variation of the temperature

range over the course of a week, in some regions, does not line up with any natural cycles that can be detected.

The average temperature range for the weekends (Saturday, Sunday, and Monday) varied from the average temperature range for weekdays (Tuesday, Wednesday, Thursday, and Friday)! Fluctuations in the daily range can be caused by natural factors; storm systems moving across an area, for example, can cause such fluctuation—but there are no natural factors known to fall consistently on certain days of the week.

The precise cause of this extraordinary pattern is not clear. However, contend the researchers (Piers M. deF. Forster and Susan Solomon), the only possible explanation for this weekend/weekday disparity is *human* activity and the atmospheric pollutants such activity creates.

—Reported in *Proceedings of the National Academy of Sciences*, 30 September 2003

5. It is no longer open to discussion that air has weight. It is common knowledge that a balloon is heavier when inflated than when empty, which is proof enough. For if the air were light, the more the balloon was inflated, the lighter the whole would be, since there would be more air in it. But since, on the contrary, when more air is put in, the whole becomes heavier, it follows that each part has a weight of its own, and consequently that the air has weight.

—Blaise Pascal, *Treatise on the Weight of the Mass of the Air*, 1653

## E. THE METHOD OF CONCOMITANT VARIATION

The four methods discussed so far are all *eliminative* in nature. By eliminating some possible cause or causes of a given phenomenon, they support some other causal account hypothesized. The method of agreement eliminates as possible causes those circumstances in whose absence the phenomenon can nevertheless occur; the method of difference permits the elimination of some possible causes by removing an antecedent factor shown to be critical; the joint method is eliminative in both of these ways; and the method of residues seeks to eliminate as possible causes those circumstances whose effects have already been established by previous inductions.

However, in many situations, no one of these methods is applicable, because they involve circumstances that cannot possibly be eliminated. This is often the case in economics, in physics, in medicine, and wherever the general increase or decrease of one factor results in a concomitant increase or decrease of another—the complete elimination of either factor not being feasible.



John Stuart Mill wrote:

Whatever phenomenon varies in any manner whenever another phenomenon varies in some particular manner is either a cause or an effect of that phenomenon or is connected with it through some fact of causation.

*Concomitant variation* is critical to the study of the causal impact of certain foods, for example. We cannot eliminate disease, no matter the diet; we can rarely eliminate foods of certain kinds from the diets of large populations. But we can note the effect of increasing or decreasing the intake of certain foods on the frequency of certain diseases in specified populations. One investigation of this kind examined the frequency of heart attacks compared to the frequency with which fish had been eaten by those in the study. The inductive conclusion was striking: Eating one fish meal a week reduced the risk of heart attack by 50 percent; eating just two fish meals a month reduced the risk of heart attack by 30 percent. Within some limits there appears to be a marked concomitant variation between cardiac arrests and the use of fish in the diet.<sup>12</sup>

Using plus and minus signs to indicate the greater or lesser degree to which a varying phenomenon is present in a given situation, the **method of concomitant variation** can be schematized as:

$$A B C - x y z.$$

$$A+ B C - x+ y z.$$


---

Therefore A and x are causally connected.

This method is very widely used. A farmer establishes that there is a causal connection between the application of fertilizer to the soil and the size of the crop by applying different amounts to different parts of a field, then noting the concomitant variation between the amounts of the additive and the yield. A merchant seeks to verify the efficacy of advertising of different kinds by running varied advertisements at varying intervals, then noting the concomitant increase or decrease of business during some of those periods.

Concomitant variation is exemplified in the search for the causes of divorce and of other important decisions among families. Of course the cause of any particular divorce will lie in the special circumstances of that marriage and that family, but there are conditions that tend generally to contribute to the breakup of families, and concomitant variation is useful in learning what these are. Analysis of data from the U.S. Census Bureau reveals that, in every decade since the 1940s, and in every region of the country, couples who were the parents of only girls divorced more often than couples who were the parents of only boys. It happened among whites and among blacks, among those with only high school diplomas and among those with college degrees. Parents with an only child who is a girl are 6 percent more likely to split up

than parents of a single boy. The gap rises to 8 percent for parents of two girls versus parents of two boys, 10 percent for families with three girls, and 13 percent if there are four girls. Thousands upon thousands of U.S. divorces appear to stem partly from the number of girls in the family.

The age-old favoring of boys, overt and common in China, India, and other developing countries, is more subtle in the United States, but it remains a widespread factor in the dynamics of U.S. family life. Parents invest more in their sons, spending, when their families include a boy, an average of an additional \$600 a year on housing. Fathers increase their workweeks after the birth of the first family child of either sex—but increase it by more than two hours if the child is a boy, less than one hour if it is a girl. These patterns of concomitant variation make it plain that parents have a preference for boys—a preference that will have increasingly important consequences when the technology for the selection of the sex of a baby, already known and reliable, becomes more widely available.<sup>13</sup>

When the increase of one phenomenon parallels the increase of another, we say that the phenomena vary *directly* with each other. However, the method permits the use of variation “in any manner,” and we may also infer a causal connection when the phenomena vary *inversely*—the increase of one leading to the decrease of another. Thus economists often say that, other things remaining roughly stable, in an unregulated market an increase of the supply of some good (say, crude oil) will result in a concomitant decrease in its price. That relation does appear to be genuinely concomitant: When international tension threatens to reduce the available supply of crude oil, we note that the price of the oil almost invariably rises.

Some concomitant variations are entirely coincidental, of course. Care must be taken not to infer a causal connection from patterns of occurrence that are wholly fortuitous. But some variations that appear to be coincidental, or are otherwise puzzling, may have an obscure causal explanation. It has been shown that there is a high correlation between the number of storks found nesting in English villages and the number of babies born in each of those villages—the more storks, the more babies. Surely it is not possible that . . . No, it's not. Villages with high birth rates have more newly married couples, and therefore have more newly constructed houses. Storks, it turns out, prefer to nest beside chimneys that have not been used previously by other storks.<sup>14</sup> Tracing the causal chains of phenomena that vary concomitantly, we may find links in common, which is what Mill meant when he said that the phenomena may be “connected . . . through some fact of causation.”

Because the method of concomitant variation permits us to adduce, as evidence, changes in the *degree* to which circumstances and phenomena are present, it greatly strengthens our set of inductive techniques. It is a *quantitative*

method of inductive inference, those earlier discussed being essentially qualitative. The use of concomitant variation therefore presupposes the existence of some method of measuring or estimating, even if only roughly, the degrees to which phenomena vary.

## EXERCISES

Analyze each of the following arguments in terms of the variation of the “phenomena” to show how they follow the pattern of the method of concomitant variation.

1. The notion that poverty and mental illness are intertwined is not new—but finding evidence that one begets the other has often proved difficult. New research, which coincided with the opening of a new gambling casino on an Indian reservation, appears to strengthen that link, strongly suggesting that lifting children out of poverty (as casino income did in many cases) tends to diminish some (but not all) psychiatric symptoms.

A study published in the *Journal of the American Medical Association* in October 2003 tracked 1420 children, ages 9 to 13, in rural North Carolina, very many of whom lived on a Cherokee Indian reservation. During the study a casino that had been opened on the reservation began distributing some of its profits to tribal families, the payments reaching about \$6000 per year by 2001. The researchers found that the rate of psychiatric symptoms among the children who had risen from poverty dropped steadily; those children were less inclined to temper tantrums, stealing, bullying, and vandalism—common symptoms of oppositional defiant disorders.

Children whose families rose above the poverty threshold showed a 40 percent decrease in behavioral symptoms. The rate of such behaviors, after four years, dropped to the same levels found among children whose families had never been poor. But the casino payments had no effect on children whose families were nevertheless unable to rise from poverty, or on those children whose families had not been poor to begin with.

The economic change had a significant effect on only a fraction of the children followed. This, it was hypothesized, was a consequence of the fact that, although all the families that received the payment received the same amount of money, the payments resulted in lifting only 14 percent of those families above the poverty line which, in 2002, was \$14,348 for a family of three. The study suggests, said Dr. Arline Geronimus, of the University of Michigan, that poverty puts stress on families, which can increase the likelihood that children will develop behavioral problems.

2. In Finland, heart attacks occur more frequently in the eastern part of the country than in the western and southern parts. Researchers seeking to explain these differences concluded that they "cannot be explained by individual lifestyle or by genetic factors." How, then, can they be explained? A study led by Dr. Anne Kousa, of the Geological Survey of Finland, examined heart attacks that occurred in 18,946 men, ages 35 to 74, in three different years. The researchers then correlated the incidence of heart attack in these populations with the level of water hardness—as measured by the presence of minerals in the water—in their communities. The study found that the degree of water hardness correlated directly with a lowered risk of heart attack. Drinking water rich in minerals appears to play a role in reducing heart disease.

—*Journal of Epidemiology and Community Health*, January 2004

3. When it comes to love, sex, and friendship, do birds of a feather flock together? Or is it more important that opposites attract? Dr. Claus Wedekind, of Bern University in Switzerland, hypothesized that body odor might signal that its owner had desirable immune genes—called MHC genes—that would help offspring to fight off diseases. He devised an experiment to see if human body odor correlated with MHC genes and if people could tell.

He and his team collected DNA samples for 49 female and 44 male university students. He asked the men to wear cotton T-shirts on two successive nights, to keep the shirt in a plastic bag, to use perfume-free detergents and soaps, and to avoid smelly rooms, smell-producing foods, and activities like smoking and sex that create odors. Meanwhile, the women were given a nasal spray to protect their nasal membranes from infection, and each received a copy of the Patrick Susskind novel *Perfume* to make them more conscious of odors.

When the T-shirts were collected, the women were asked to give ratings, for intensity, pleasantness, and sexiness, to three T-shirts from men with similar MHC genes, and three from men with dissimilar MHC genes, not knowing which was which.

Women who were dissimilar to a particular male's MHC perceived his odor as more pleasant than did women whose MHC was similar to the test man. Odors of men with dissimilar MHC reminded the women of their own mates or former mates twice as often as did the odors of men with similar MHC.

However, if a woman was taking oral contraceptives, which partly mimic pregnancy, this predilection was reversed, and they gave higher rating to men with similar MHC. "The Pill effect really surprised me," said Dr. Wedekind.

—*Proceedings of the Royal Society of London*, 1995

4. Stanley Coren sought to plumb the connections between sleeplessness and accidents. To do that he focused on the yearly shift to daylight time in eastern North America when (because clocks are moved forward one hour) most people lose an hour of sleep. He compared the number of accidents then with the number on normal days, and found that on the day after the time change, in Canada, there was an 8 percent increase in accidents. Then, examining the day after the return to standard time, when people gain an hour of sleep, he found a corresponding decrease in accidents. "What we're looking at," says the Director of the Human Chronobiology Laboratory at the University of Pittsburgh, commenting upon Soren's results, "is national jet lag."

—S. Coren, *Sleep Thieves* (New York: The Free Press, 1996)

5. Prof. Kathleen Vohs reports that two groups of college students were asked to read out loud from "a boring book on the biographies of scientists." One of the groups was obliged to wear fake expressions of delight and interest, while the other group was allowed to read the same texts naturally. Each group was afterward given a sum of money to spend on an assortment of goods, or to save. Those who had been faking delight spent 62 percent more than those who had not. Similarly, a group of students writing down their thoughts without restraint spent very much less than a similar group obliged while writing to avoid all thoughts about white bears. The more self-restraint that a person expends to control one impulse, it appears, the less self-restraint is available to control others.

—Reported in *The Journal of Consumer Research*, March 2007

6. Potassium in the urine is known to reflect potassium intake from the diet. At the Prosserman Center for Health Research in Toronto, Dr. Andrew Mente and colleagues analyzed urinary potassium as a useful clinical marker of a healthy diet. They collected urine samples from hundreds of patients and separately calculated the quality of their diets. The results were striking: as urinary potassium increased there was a steady and significant increase in diet quality score, as well as a steady decrease in body mass, blood pressure and heart rate. "This urinary marker," said Dr. Mente, "is a simple, objective, universally available measure of diet quality."

—*Urology/Nephrology News*, 20 November 2006

7. Whenever the U.S. says things that make a military conflict with Iran seem more likely, the price of oil rises, strengthening Iran's regime rather than weakening it. The more we talk about curbing Iranian



power, the more difficult it gets. . . . So cooling down the martial rhetoric, even if we plan to take military action eventually, would likely bring oil prices down, making Iran weaker. . . . Lower oil prices won't, by themselves, topple the mullahs in Iran. But it's significant that, historically, when oil prices have been low, Iranian reformers have been ascendant and radicals relatively subdued, and vice versa when prices have been high. Talking tough may look like a good way of demonstrating U.S. resolve, but when tough talk makes our opponent richer and stronger we may accomplish more by saying less.

—James Surowiecki, "Troubled Waters over Oil,"  
*The New Yorker*, 19 February 2007

## OVERVIEW

### Five Methods of Inductive Inference

1. **The method of agreement.** The one factor or circumstance that is *common* to all the cases of the phenomenon under investigation is likely to be the cause (or effect) of that phenomenon.
2. **The method of difference.** The one factor or circumstance whose absence or presence *distinguishes* all cases in which the phenomenon under investigation occurs from those cases in which it does not occur, is likely to be the cause, or part of the cause, of that phenomenon.
3. **The joint method of agreement and difference.** Although perhaps not a separate method, *the combination*, in the same investigation, of *the method of agreement and the method of difference* gives substantial probability to the inductive conclusion.
4. **The method of residues.** When some portion of the phenomenon under examination is known to be the consequence of well-understood antecedent circumstances, we may infer that *the remainder of that phenomenon is the effect of the remaining antecedents*.
5. **The method of concomitant variation.** *When the variations in one phenomenon are highly correlated with the variations in another phenomenon*, one of the two is likely to be the cause of the other, or they may be related as the products of some third factor causing both.

These are the inductive methods frequently called **Mill's methods**, most commonly used by scientists in their investigation of causal laws.

## 12.5 Limitations of Inductive Techniques

---

What do the methods explained in the preceding sections actually do for us? John Stuart Mill believed that they were instruments with which we may *discover* causal connections, and also that they were canons with which causal connections may be *proved*. On both counts he overestimated their power. Inductive techniques are indeed of very great importance, but their role in science is more limited than Mill supposed.

One substantial difficulty arises from the fact that, in formulating these methods, Mill made the assumption that one can identify cases "having *only* one circumstance in common" or other cases "having *every* circumstance in common save one." But these expressions must not be taken literally; any two objects will have many circumstances in common however different they may appear; and no two things can ever differ in only one respect—one will be farther to the north, one will be closer to the sun, and so on. Nor could we even examine all possible circumstances to determine if they differ in only one way. What the scientist has in mind as he applies these techniques are not all circumstances, but the sets of *relevant* circumstances—whether there is only one relevant circumstance in common, or all relevant circumstances save one in common. That is, we apply the methods to the circumstances that have some bearing on the causal connection in question.

Which are those circumstances? We cannot learn which factors are relevant using the methods alone. In order to use the methods we must come to the context in which they are to be applied with some analysis of causal factors already in mind. The caricature of the "scientific drinker" illustrates this difficulty: He drinks Scotch and soda one night, bourbon and soda the next night, and on the following nights brandy and soda, then rum and soda, then gin and soda. What is the *cause* of his intoxication? Repeatedly inebriated, he swears never to touch soda again!

This scientific drinker did apply the method of agreement in accordance with the rules—but his doing so was to no avail because the factors that really are relevant in those antecedent circumstances had not been identified and therefore could not be manipulated. Had *alcohol* been specified as one of the factors common to all the cases, it would have been possible to eliminate soda very quickly, of course, using the method of difference.

The heroic investigation of the causes of yellow fever, discussed earlier in connection with the method of difference, confirmed the conclusion that the fever is spread by the bite of an infected mosquito. We know that *now*, just as we know now that it is alcohol and not soda that causes drunkenness. But the yellow fever experiments required insight and imagination as well as courage;

the notion that the fever was spread by mosquitoes was originally thought to be silly, or absurd, or was not thought of at all. Circumstances in the real world do not come wearing tags marked “relevant” or “irrelevant.” The testing of mosquito bites as cause required some earlier sorting of possibly relevant factors, to which the inductive methods might then be applied. With that prior analysis in hand, the methods can prove exceedingly helpful—but the methods by themselves, without some hypotheses in the background, are not *sufficient* instruments for scientific discovery.

Nor can the methods by themselves constitute rules for *proof*. Their application proceeds always on the basis of some antecedent hypotheses about causal factors, as noted just above, and because all circumstances cannot have been considered, attention will be confined to those believed to be the possible causes in question. However, this judgment regarding which circumstances are to be investigated may prove to have been in error. Medical scientists, for a very long time, did not consider dirty hands even as *possible* agents of infection, and so could not identify such dirtiness as the cause of disease.\* Investigation is stymied when the investigators fail to break down the circumstances before them into the appropriate elements, elements that cannot be known in advance. Because the analyses presupposed by the application of the methods may be incorrect, or inadequate, the inferences based on those analyses may also prove to be mistaken. This dependence of induction on the merit of the underlying hypotheses shows that inductive techniques cannot by themselves provide the proof of causation that Mill had hoped for.

Yet another problem should be borne in mind: The application of inductive methods always depends on *observed* correlations, and even when the observations have been made accurately, they may be incomplete and therefore deceptive. The greater the number of observations, the greater is the likelihood that the correlation we observe is the manifestation of a genuine causal law—but no matter how great that number is, we cannot infer with certainty a causal connection among instances that have not yet been observed.

---

\*The failure of physicians to wash their hands (because they did not understand how infectious diseases were spread) resulted in untold misery and uncounted deaths over centuries, especially from puerperal (or childbed) fever, which was carried on the hands of doctors from mother to mother, until the proof of that disastrous causal connection was given by the Hungarian physician Ignac Semmelweis, in the middle of the nineteenth century. See Sherwin B. Nuland, *The Doctors' Plague* (New York: W. W. Norton, 2003).

These limitations illuminate once again the great gulf between deduction and induction. A valid deductive inference constitutes a proof, or demonstration, but every inductive inference is, at best, highly probable and never demonstrative. Therefore Mill's claim that his canons are "methods of proof" must be rejected, along with the claim that they are "the methods of discovery."

Nevertheless, the techniques explained in this chapter are central in much of science and are very powerful. Because it is impossible for investigators to take *all* circumstances into account, the application of the methods must always suppose one or more causal *hypotheses* about the circumstances under investigation. Being unsure which factor(s) are the cause(s) of the phenomenon under investigation, we often formulate alternative hypotheses and subject each to testing. What the five methods of induction, being mainly eliminative in nature, enable us to determine is this: *If* some specified analysis of the antecedent circumstances is correct, one of these factors cannot be (or must be) the cause (or part of the cause) of the phenomenon in question. This may be deduced, and the deduction may be valid, but the soundness of that argument will always depend on the correctness of the antecedent analysis that had been supposed.

The methods of induction are splendid, but they can yield reliable results only when the hypothesis that they seek to confirm (or falsify) does identify correctly the circumstances that are causally relevant. The methods permit the *deduction* of those results only when that hypothesis has been assumed as a *premise* in the argument. The nature of the power these methods give us may now be seen. They are not paths for discovery; they are not rules for proof. *They are instruments for testing hypotheses.* The statements of these inductive techniques, taken together, describe the general method of controlled experiment, which is a common and indispensable tool in all of modern science.

So important is the role of hypotheses in systematic empirical investigations that the enterprise of devising and testing hypotheses may be regarded as *the* method of science—to which we turn in the next chapter.

## EXERCISES

Analyze each of the following investigations, or arguments, and indicate which of the methods of causal reasoning—Mill's methods—are being used in each of them.

1. Teens who lose their virginity earlier than their peers are more likely to shoplift, destroy property, or sell drugs than their virgin counterparts, according to a recent national study of 7000 teenagers. Those who had sex early were 20 percent more likely to engage in delinquent acts one



year later compared to those whose first sexual experience occurred at the average age for their school. Those who waited longer than average to have sex had delinquency rates 50 percent lower a year later compared to average teens. Waiting appears to have a protective effect. "We're not finding that sex itself leads to delinquency; sex itself is not always a problem behavior," writes co-author and Ohio State sociologist Stacy Armour. However, "the timing of sexual initiation does matter. Kids go off on a different trajectory if they're having sex early."

—Reported in *The Journal of Youth and Adolescence*, February 2007

2. Strong evidence has been presented that a diet low in folic acid [a trace vitamin in the B complex] during pregnancy increases the chances of giving birth to a premature baby of lower than normal birth weight. Dr. Theresa Scholl [of the University of Medicine and Dentistry of New Jersey] studied the outcomes of pregnancy for 832 women from the inner city of Camden, N.J., to determine the influence of dietary and supplementary consumption of folic acid. "We found that the women who consumed less than 240 micrograms per day of folic acid had about a two to threefold greater risk of preterm delivery and low birth weight," she said. She reported that even small increases in the women's serum folic acid concentrations by the 28th week decreased the odds of preterm delivery as well as the chance of having a baby of low birth weight. Of the 219 women in the low-folic-acid category (receiving less than 240 micrograms a day), 44 had preterm, low birth weight infants. "The risks declined in direct relationship to increased serum levels of folic acid, showing that low intake is a risk factor throughout pregnancy," Dr. Scholl concluded.

—T. O. Scholl, et al., "Dietary and Serum Folate: Their Influence on the Outcome of Pregnancy," *American Journal of Clinical Nutrition*, April 1996

3. The sequence of DNA units in the genome of humans and in that of chimpanzees is 98.8 percent identical; humans and chimps shared a joint ancestor as recently as five million years ago. Relatively few genes, therefore, must define the essence of humanity, and biologists have long supposed that if they could identify genes that have *changed* in the evolutionary advance leading from that joint ancestor, they would better understand the genetic basis of how people differ from chimpanzees, and hence what makes humans human.

This project received a significant boost in 2001 when a large London family with barely intelligible speech was found to have



mutations in a gene called FOXP2. Chimpanzees also have an FOXP2 gene, but theirs is significantly different from ours. The human version shows signs of accelerated evolutionary change in the last 100,000 years, which suggests that the gene acquired a new function that helped to make human speech possible.

—Reported by Dr. Michelle Cargill of Celera Diagnostics, Alameda, CA, and Dr. Andrew Clark, of Cornell, in *Science*, 11 December 2003

4. A simple, inexpensive and surprisingly powerful combination of treatments that all but wiped out malaria in a group of HIV-positive children in a recent study in Uganda was described at a very recent medical conference in Los Angeles. The combination—taking one inexpensive antibiotic pill each day and sleeping under an insecticide-treated mosquito net—*reduced the incidence of malaria by 97 percent* compared with a control group. The study, conducted by Dr. Anne Gasasira of Makerere University in Kampala, Uganda, found that among 561 healthy children who were not HIV-infected and who did not take the antibiotic and sleep under bed nets, there were 356 episodes of malaria. This compared with 4 episodes among 300 children who were known to be HIV-infected and received both treatments. “The findings were shockingly dramatic,” said Dr. Elaine Abrams, a professor of pediatrics and epidemiology at Columbia University.

—Reported at the 14th Conference on Retroviruses and Opportunistic Infections, Los Angeles, 28 February 2007

5. Some theories arise from anecdotal evidence that is difficult to confirm. In *The Left-Hander Syndrome* (New York: Bantam Books, 1992), Stanley Coren sought to evaluate the common belief that left-handed persons die sooner than right-handers. But death certificates or other public records very rarely mention the hand preferred by the deceased. What could serve as a reliable data source with which that hypothesis could be tested? Coren searched baseball records, noting which hand baseball pitchers threw with, and then recording their ages at death. Right-handed pitchers, he found, lived on average nine months longer than lefties. Then, in a follow-up study, he and a colleague telephoned the relatives of people named on death certificates in two California counties, to ask which hand the deceased favored. Right-handed people (that study found) lived an average of nine years longer than lefties.

6. It has long been recognized that taller adults hold jobs of higher status and, on average, earn more than other workers. A large number of hypotheses have been put forward to explain the association between height and earnings. In developed countries, researchers have emphasized factors such as self esteem, social dominance, and discrimination. In this paper, we offer a simpler explanation: On average, taller people earn more because they are smarter. As early as age 3—before schooling has had a chance to play a role—and throughout childhood, taller children perform significantly better on cognitive tests. The correlation between height in childhood and adulthood is approximately 0.7 for both men and women, so that tall children are much more likely to become tall adults. As adults, taller individuals are more likely to select into higher paying occupations that require more advanced verbal and numerical skills and greater intelligence, for which they earn handsome returns. Using four data sets from the US and the UK, we find that the height premium in adult earnings can be explained by childhood scores on cognitive tests. Furthermore, we show that taller adults select into occupations that have higher cognitive skill requirements and lower physical skill demands.

—Anne Case and Christina Paxson, "Stature and Status: Height, Ability, and Labor Market Outcomes," *National Bureau of Economic Research*, Working Paper No. 12466, August 2006

7. Does the position of the arm, when blood pressure is being checked, make any difference? Researchers at the University of California at San Diego, using automated cuffs, took six readings from one hundred emergency room patients whose problems did not involve their circulatory systems. Their blood pressure was measured standing, sitting, and lying down; in each position it was measured with the arm straight out from the body and with the arm held at the side. They found that the position of the arm had a bigger effect on the readings than the position of the body. When the arm was parallel to the body readings were higher by as much as 14 millimeters of mercury. Dr. David A. Guss, one of the authors of the study, said that no single position was more accurate, "the most important thing is to use a consistent position from measurement to measurement."

—From the *Annals of Internal Medicine*, reported in  
*The New York Times*, 6 January 2004

8. Near the end of the Middle Ages, a few theologians (the "scientists" of that time) persuaded a king of France to give them permission for an

experiment that had been forbidden by the Roman Catholic Church. They were allowed to weigh the soul of a criminal by measuring him both before and after his hanging. As usually happens with academics, they came up with a definite result: the soul weighed about an ounce and a half.

—John Lukacs, "Atom Smasher Is Super Nonsense,"  
*The New York Times*, 17 June 1993

9. Undoubtedly the outstanding point of departure of industrial social psychology was the series of studies performed in the Hawthorne plant of the Western Electric Company, starting in 1927. These were conducted by three Harvard professors, Elton Mayo, F. J. Roethlisberger, and T. N. Whitehead, and by W. J. Dickson of Western Electric. The original aim of the studies was to obtain concrete data on the effects of illumination, temperature, rest periods, hours of work, wage rate, etc., upon production. A group of six girls, average workers, were chosen for the experiment; their task was the assembly of telephone relays. Almost from the beginning, unexpected results appeared: The production rate kept going up whether rest periods and hours were increased or decreased! In each experimental period, whatever its conditions, output was higher than in the preceding one. The answer seemed to lie in a number of subtle social factors.

. . . As Homans summarizes it, the increase in the girls' output rate "could not be related to any change in their conditions of work, whether experimentally induced or not. It could, however, be related to what can only be spoken of as the development of an organized social group in a peculiar and effective relation with its supervisors."

—S. Stansfeld Sargent and Robert C. Williamson, *Social Psychology*, 1966

10. Does noise have an adverse effect on those subjected involuntarily to it? When the airport at Munich, Germany, moved, researchers from the University of Hamburg, the University of Gavle in Sweden, and Cornell University took that rare opportunity to conduct a prospective study on the effects of noise, measuring the performance of students near the old airport and near the new one, before and after the move. The reading skills of students in both groups were tested, along with short-term and long-term memory, as reported in the journal *Psychological Science*, in October 2002. After the move, improvements in memory and reading were found among students near the old airport, while among students living near the new airport, reading skills and memory performance declined.

High levels of noise do interfere with learning and development, those researchers concluded—but the brighter side of their findings was this: Most of the learning damage done by noise appeared to reverse itself when the noise was removed.

11. The mood changes that many people experience during the shorter days of winter have a physiological basis in the brain, according to a study reported in the British medical journal, *The Lancet*, in January 2003. One hundred healthy volunteers, ages 18 to 79, allowed researchers to draw blood samples, at different times of the year, from their jugular veins, to get blood as close to the brain as possible. The researchers then correlated levels of brain chemicals, especially serotonin, with the weather data—temperature, air pressure, rainfall, and sunlight—at the times of blood collection. Only sunlight had causal impact; serotonin levels were found to be lowest in the three months of winter, but varied depending on the brightness of the day. “Our findings [the researchers wrote] are further evidence for the notion that changes in release of serotonin by the brain underlie mood seasonality and seasonal affective disorder.”
12. Prof. Norbert Schwartz, of the University of Michigan, conducted the following experiment. He tested the attitudes of people who had just used a University of Michigan copying machine in which, for some subjects, he had planted a dime which they found, while for others there was no windfall dime. After using the copier, subjects were asked how happy they were about life. Those who had found a dime were consistently more upbeat about “their lives as a whole,” and about the economy and many other matters. “We found,” said Prof. Schwartz, “that a dime can make you happy for about twenty minutes. Then the mood wears off.”
 

—N. Schwartz, *Well Being: Foundations of Hedonic Psychology*  
(New York: Russell Sage Foundation, 1999)
13. The largest and longest-running study of American child care has found that keeping a preschooler in a day care center for a year or more increased the likelihood that the child would become disruptive in class—and that this effect persisted through the sixth grade. Every year spent in such centers for at least 10 hours per week was associated with a 1 per cent higher score on a standardized assessment of problem behaviors completed by teachers. Parents’ guidance, and their genes, had the strongest influence on how children behaved—but this finding about the impact of day care centers held up regardless of the

child's sex, or family income, and regardless of the quality of the day care center.

—National Institute of Child Health and Human Development,  
"Early Child Care and Youth Development," 26 March 2007

14. Speed kills. A report from the Insurance Institute for Highway Safety, issued in November of 2003, concluded that increased speed limits on Interstate highways led to nearly 1,900 additional deaths in 22 states from 1996 to 1999. The report is based, oddly, on a study by the Transport Safety Authority of New Zealand, working in the United States, which showed that, when the Federal cap on speed limits was placed at 65 mph the number of deaths on U.S. highways decreased. But almost immediately after the repeal of that Federal cap on speed limits the number of deaths in the states that did not retain the 65 mph limit increased markedly, while the number of deaths in those states that retained the 65 mph limit did not increase. Drivers in states with higher speed limits, the study showed, drive faster, and where the driving is faster the number of traffic fatalities goes up.

—"Study Links Higher Speed Limits to Deaths"  
*The New York Times*, 24 November 2003

15. A 16-year study followed 8,867 non-smoking male professionals with normal body weight who participated in vigorous daily exercise and ate a healthy diet. Those who drank one-half to two normal servings of wine, beer, or hard liquor a day had a 41 to 62 percent reduction of heart attack risk compared with those who drank no alcohol at all. It seems clear that in moderate quantities alcoholic drinks reduce the likelihood of heart attack. This effect is found not only in those with heart disease. The lead author of the study writes: "Even in the lowest risk people, we still find a lower risk associated with moderate drinking."

—Kenneth Mukamal, "Alcohol Consumption and Risk for  
Coronary Heart Disease in Men with Healthy Lifestyles,"  
*Archives of Internal Medicine*, 23 October 2006

16. For heart patients, "noetic" intervention, like prayer, and therapy relying on music, imagery and touch (MIT), is defined as "an intangible healing influence brought about without the use of a drug, device, or surgical procedure." 748 patients with coronary heart disease who were to undergo percutaneous coronary intervention (a type of stenting procedure), or elective cardiac catheterization, were enrolled



at one of nine study sites between 1999 and 2002. To test the efficacy of noetic intervention, patients were randomized into four groups: one group (189 patients) received both offsite intercessory prayer and MIT therapy; a second group (182 patients) received intercessory prayer only; a third group (185 patients) received MIT therapy only; the fourth group (192 patients) received neither the intercessory prayer nor the MIT therapy. The interventional heart procedures were conducted according to each institution's standards practice, with a six-month period of follow-up. The prayer portion was double blinded, meaning that the patients and their care team did not know which patients were receiving intercessory prayer. The prayer groups for the study were located throughout the world and included Buddhist, Muslim, Jewish and many Christian denominations. 89 percent of the patients in this study also knew of someone praying for them outside of the study protocol.

As reported by the Duke University Medical Center, the researchers found no significant difference among the four treatment groups. Distant prayer and the bedside use of music, imagery and touch did not have a significant effect upon the primary clinical outcome of these patients undergoing medical interventions.

—"First Multicenter Trial of Intercessory Prayer," *The Lancet*, 16 July 2005

17. The impulse to share does not come naturally to one who is thinking about money. Psychologists found that subconscious reminders of money prompted people to become more independent in their work, and less likely to seek help from others or to provide it. In one experiment 52 undergraduates unscrambled sets of jumbled phrases; one group untangled phrases that were often about money, like "high salary paying," while another solved word puzzles that did not refer to money. Researchers then had the students work on a difficult abstract puzzle and offered to give help if they wanted it. Those who had been thinking about money worked on the problem by themselves an average of more than 70% longer than the others. Students "primed" to have money on their minds, while clearly self-reliant, were less likely than peers who had not been so primed to lend assistance, twice as slow to help another confused student, and about twice as stingy when asked to donate money to help needy students.

—Kathleen Vohs, Nicole Mead, and Miranda Goode, "The Psychological Consequences of Money," *Science*, 17 November 2006

---

## SUMMARY

---

In this chapter we have examined the concept of cause, the nature of causal connections, and the methods used to establish causal laws.

In Section 12.1, we examined various meanings of "cause."

In Section 12.2, we explained the supposition of the uniformity of nature, and the generality of causal laws.

In Section 12.3, we discussed induction by simple enumeration.

In Section 12.4, we recounted and illustrated the principal techniques of inductive inference, called Mill's methods, explaining their essentially eliminative nature. These five methods are

1. The method of agreement
2. The method of difference
3. The joint method of agreement and difference
4. The method of residues
5. The method of concomitant variation

In Section 12.5, we explained the limitations and the strengths of these inductive techniques, concluding that, although they cannot do all that John Stuart Mill had claimed for them, they are profoundly important as the intellectual instruments with which scientific hypotheses are confirmed or disconfirmed.

---

## End Notes

---

<sup>1</sup>David Hume, *An Enquiry Concerning Human Understanding* (1748), sec. IV.

<sup>2</sup>J. Dao, "Coroner in Cincinnati Rules Man's Struggle Led to Death," *The New York Times*, 4 December 2003.

<sup>3</sup>Reported by Connor O'Shea of the Duke University Medical Center, at the meetings of the European Society of Cardiology in August 2000.

<sup>4</sup>See Zachariah Chafee, Jr. *Three Human Rights in the Constitution of 1787* (1952).

<sup>5</sup>Dr. Steven Grant, quoted in *Science News*, 27 January 2007.

<sup>6</sup>Here quoting Dr. Nora Volkow, director of the National Institute of Drug Abuse, quoted by Benedict Carey, "In Clue to Addictive Behavior, A Brain Injury Halts Smoking," *The New York Times*, 26 January 2007.

<sup>7</sup>Robert Sapolsky, "Testosterone Rules," *Discover*, March 1997.

<sup>8</sup>D. N. Cook *et al.*, "Requirement of MIP-1 *alpha* for an Inflammatory Response to Viral Infection," *Science*, 15 September 1995.

<sup>9</sup>Paul Henle and William K. Frankena, *Exercises in Elementary Logic* (1940).

<sup>10</sup>A. Werzberger *et al.*, "A Controlled Trial of a Formalin-Inactivated Hepatitis A Vaccine in Healthy Children," *The New England Journal of Medicine*, 13 August 1992.

<sup>11</sup>Edward Arthur Fath, *The Elements of Astronomy* (New York: McGraw-Hill, 1926), p. 170.

<sup>12</sup>D. S. Siscovick *et al.*, "Dietary Intake and Cell Membrane Levels of Long-Chain *n*-3 Polyunsaturated Fatty Acids and the Risk of Primary Cardiac Arrest," *Journal of the American Medical Association*, 1 November 1995.

<sup>13</sup>The source of these data is the U.S. Census Bureau; the analysts are Gordon B. Dahl of the University of Rochester and Enrico Moretti of the University of California at Los Angeles, reporting online in *Slate*, in October 2003.

<sup>14</sup>J. L. Casti, *Searching for Certainty* (New York: William Morrow, 1991).

## SUMMARY

---

In this chapter we have examined the concept of cause, the nature of causal connections, and the methods used to establish causal laws.

In Section 12.1, we examined various meanings of "cause."

In Section 12.2, we explained the supposition of the uniformity of nature, and the generality of causal laws.

In Section 12.3, we discussed induction by simple enumeration.

In Section 12.4, we recounted and illustrated the principal techniques of inductive inference, called Mill's methods, explaining their essentially eliminative nature. These five methods are

1. The method of agreement
2. The method of difference
3. The joint method of agreement and difference
4. The method of residues
5. The method of concomitant variation

In Section 12.5, we explained the limitations and the strengths of these inductive techniques, concluding that, although they cannot do all that John Stuart Mill had claimed for them, they are profoundly important as the intellectual instruments with which scientific hypotheses are confirmed or disconfirmed.

## End Notes

---

<sup>1</sup>David Hume, *An Enquiry Concerning Human Understanding* (1748), sec. IV.

<sup>2</sup>J. Dao, "Coroner in Cincinnati Rules Man's Struggle Led to Death," *The New York Times*, 4 December 2003.

<sup>3</sup>Reported by Connor O'Shea of the Duke University Medical Center, at the meetings of the European Society of Cardiology in August 2000.

<sup>4</sup>See Zachariah Chafee, Jr. *Three Human Rights in the Constitution of 1787* (1952).

<sup>5</sup>Dr. Steven Grant, quoted in *Science News*, 27 January 2007.

<sup>6</sup>Here quoting Dr. Nora Volkow, director of the National Institute of Drug Abuse, quoted by Benedict Carey, "In Clue to Addictive Behavior, A Brain Injury Halts Smoking," *The New York Times*, 26 January 2007.

<sup>7</sup>Robert Sapolsky, "Testosterone Rules," *Discover*, March 1997.

<sup>8</sup>D. N. Cook *et al.*, "Requirement of MIP-1 *alpha* for an Inflammatory Response to Viral Infection," *Science*, 15 September 1995.

<sup>9</sup>Paul Henle and William K. Frankena, *Exercises in Elementary Logic* (1940).

<sup>10</sup>A. Werzberger *et al.*, "A Controlled Trial of a Formalin-Inactivated Hepatitis A Vaccine in Healthy Children," *The New England Journal of Medicine*, 13 August 1992.

<sup>11</sup>Edward Arthur Fath, *The Elements of Astronomy* (New York: McGraw-Hill, 1926), p. 170.

<sup>12</sup>D. S. Siscovick *et al.*, "Dietary Intake and Cell Membrane Levels of Long-Chain *n*-3 Polyunsaturated Fatty Acids and the Risk of Primary Cardiac Arrest," *Journal of the American Medical Association*, 1 November 1995.

<sup>13</sup>The source of these data is the U.S. Census Bureau; the analysts are Gordon B. Dahl of the University of Rochester and Enrico Moretti of the University of California at Los Angeles, reporting online in *Slate*, in October 2003.

<sup>14</sup>J. L. Casti, *Searching for Certainty* (New York: William Morrow, 1991).