11. SCIENTIFIC REASONING

In Part 4, we turn to applications of the fundamental concepts covered in earlier chapters. In particular, the next four chapters examine some of the methods and forms of reasoning found in scientific contexts (Chapter 11), the forms of reasoning used for deliberating over moral issues (Chapter 12), strategies for reasoning in criminal and civil law (Chapter 13), and effective ways of responding to arguments by other people (Chapter 14). Each of these contexts requires us to argue with a fairly well-defined purpose in mind, and these purposes may be described in broad terms. One important purpose of scientific reasoning is to expand our knowledge of the world by answering questions about *why* things happen as they do. The purpose of moral reasoning is to formulate reasonable moral judgments about what is right, wrong, good, bad, fair, and unfair. The purpose of legal reasoning is to formulate judgments that decide legal contests. And the purpose of arguing back is to engage the arguments of other people in a rational and critical manner. Each of these purposes presents special difficulties that must be dealt with in order to argue effectively in these contexts.

An essential task of science is to explain how and why things happen. From physical sciences (such as mechanics, chemistry, and biology) to social sciences (such as economics, sociology, and psychology), every branch of science has this task. For this reason, a great deal of effort is invested in formulating, testing, and refining the explanatory concepts that are at the heart of each scientific discipline. Accordingly, part of this chapter is devoted to some of the inductive procedures that help generate explanatory causal statements (section 11.2). Another part examines some problems connected to the application of explanatory hypotheses, in particular, deciding which ones to accept and apply when more than one is available (section 11.4). We finish the chapter with a case that will illustrate how the various methods, argument forms, and concepts that we have covered can be used in combination with each other (section 11.5). But, first, let us consider what makes a scientific concept genuinely explanatory.

11.1 CAUSATION / CORRELATION

To appreciate some important widespread features of scientific reasoning, let us return to a concept that first came up in connection with argument adequacy:

causation. In Chapter 8 we noted in passing the elusiveness of causation before we turned to the fallacies of *post hoc*, *confusing cause and effect*, and *common cause*. Now we must consider causation in a little more detail, because it is a central concept in many different kinds of scientific reasoning. Since Aristotle's time, philosophers and scientists have attempted to understand many types of phenomena in a scientific way by discovering their causes. In these cases, we explain particular events and general patterns by identifying the causal factors involved.

To begin, we must note that there are a couple of ways in which two or more events may be related. First, there is a causal relation that is important for a scientific understanding. Second, there is the temporal relation of events, i.e., their order in time. We can *observe* the order of two events in time and note whether one occurred before, after, or simultaneously with another. In other words, a statement about the temporal relation of events is empirical. But when we claim that one event *caused* another one, the connection between the two is much stronger than anything we observe directly about their order in time. When someone says that moderate levels of rainfall in June *caused* a bumper crop of corn the following August, that claim goes beyond what is strictly observable about the temporal order of these events. Similarly, to say that an explosion was *caused* by a gas leak in the presence of an open flame goes beyond the strict observation that both the gas leak and the open flame occurred simultaneously with the explosion. As both these cases illustrate, while the temporal relation is observable, the causal relation is not. This is because a scientific (or complete) causal account specifies the necessary and sufficient conditions for something to occur, and both of these conditions involve counterfactual statements (see section 1.3). Counterfactual statements are about what would have happened had the purported necessary and sufficient conditions not been satisfied. These possibilities are not directly observed; one purpose for experimental research, then, is to test whether these implicit claims about necessary and sufficient conditions can be satisfied empirically. A *scientific* causal account, which is the ideal of scientific understanding, specifies all of these conditions in terms of general principles; by contrast, a *non-scientific* (or simple) causal account specifies at least some of the conditions that are necessary and sufficient for something, many of which may be particular to the single instance in question. An inquiry into the gas-plant leak, for example, may focus on a single loose coupling and an errantly discarded cigarette as the particular conditions that constitute the simple cause of the explosion. A scientific account of one year's bumper crop of corn, by contrast, would explain the general and systematic relationships that hold between crop yields and growing conditions.

For these reasons, the concept of causation must be carefully distinguished from the concept of correlation. Two events that regularly occur at the same time or in the same sequence may be both correlated and related as cause and effect *or* they may be correlated without being in a direct causal relationship. A **CORRELATION** is observed when different events occur at the same time or occur regularly in the same sequence. With **CAUSATION**, one event (the cause) is responsible for, or brings about, another event (the effect). We can see the need for this distinction by considering one of the causal fallacies outlined earlier: common cause (see section 8.6.3). Someone might notice that a sore throat is always accompanied by sinus congestion (a correlation). On the basis of this observed correlation the sick person might fallaciously believe that the sinus congestion *causes* the sore throat. But really the sore throat and the sinus congestion are both caused by a third factor, namely, a cold virus. So while the sore throat and congestion have a common cause, neither causes the other. In this case, there is a correlation but no causal relationship between the sore throat and the congestion. At the same time, there is both a correlation and a causal relationship between the cold virus (the cause) and the congestion (an effect) and between the virus and the sore throat (another effect). If the sick person makes the effort to explore these relations, she may be able to correct her own fallacious reasoning. Suppose this person takes a decongestant that relieves the sinus congestion but does nothing for the sore throat. This "experiment" would reveal the falsity of a counterfactual implicit in the original causal hypothesis, i.e., *If the sinuses were not congested, then the throat would not be sore.* This person is now in a much better position to understand both the sore throat and sinus congestion as effects of some other cause.

Causal claims attempt to identify what is responsible for, or what brings about, a state of affairs. They answer questions that take the form *Why ...?* For example, we might ask

Why does the price of gold go up when the stock market is down?
Why do the tides alternate as they do?
Why did the plane miss the runway during landing?
Why did wheat production in Nebraska drop dramatically during the 1930s?

Each of these questions can be reworded as a question about what causal factor is responsible for the event in question:

What is responsible for the rising price of gold when the stock market is down?
What is responsible for the alternation of tides?
What was responsible for the plane missing the runway during landing?
What was responsible for the drop in wheat production in Nebraska in the 1930s?

Under either formulation, each question treats the phenomenon being asked about as an effect. Also, each one is answered when the cause is specified:

> *Investors seeking a secure commodity during periods of market uncertainty cause the price of gold to rise.*
> *The mutual gravitational attraction of the moon and the earth causes the tides.*
> *Fog and pilot error caused the plane to miss the runway.*
> *Drought, soil erosion, and swarms of crop-eating insects caused a reduction in wheat production.*

In each example, the answer is a circumstance (or set of circumstances) that is both correlated with the phenomenon being asked about and responsible for it. Because the correlated circumstances usually occur prior to the event in question, they are called **ANTECEDENT CIRCUMSTANCES**.

Mere correlations identify antecedent circumstances without answering *Why ...?* questions, such as those found in the previous paragraph. Suppose the price of gold happens to go up every time polka dot ties become fashionable. In this case, we have a strong correlation between two phenomena. However, there are no good grounds to causally and directly connect trends in tie fashion and trends in the gold market. Even if the correlation has been strong in the past, we shouldn't expect it to be repeated in the future. On its own, this correlation does not answer the *Why ...?* question about the price of gold, and we have no reason to believe it has been anything more than a curious coincidence. To answer a *Why ...?* question, it isn't enough to notice what circumstances precede the phenomenon in question. We must have some reason to believe that the correlated antecedent circumstances played a role in bringing about the event.

Of course, noticing a correlation between one event and another can help us to identify the cause of some phenomenon, even if a correlation is not sufficient to establish a causal relation. Where there is *no* correlation, we will find no causal relation between two events. For example, because the phases of Venus are not correlated with patterns associated with tides, we discount these phases as a cause of the tides. But because the position of the moon with respect to the earth is correlated with the tides, we have some reason to suspect that the moon might be involved in the cause we are looking for. If a correlation holds in a wide range of instances and over an extended period of time, then we should suspect that an underlying causal relationship is at work here. Better still, if all the correlated incidents can be organized systematically, then we can use our observations to *infer* a causal relationship. The methods of reasoning described in the next section are designed to help determine whether certain correlations provide strong grounds for making causal claims.

11.2 MILL'S METHODS

The nineteenth-century philosopher John Stuart Mill formulated several ways of collecting and organizing observations so that they could be used to infer causal relationships. These inferences will always be inductive, because they require us to extrapolate from premises about observable events to conclusions about the underlying causal relationship between these events. So the conclusions we draw on the basis of these methods can never be guaranteed in the way deductive conclusions are. What Mill has provided is a set of procedures for systematically and methodically extrapolating from the observable temporal order of things to the causal order that is not directly observable. By being methodical in the ways he recommends, we can increase the probability that our conclusions about the causal order are correct.

Mill outlines five methods: (1) the Method of Agreement, (2) the Method of Difference, (3) the Joint Method of Agreement and Difference, (4) the Method of Concomitant Variations, and (5) the Method of Residue.

11.2.1 Method of Agreement

When we observe several instances of the same phenomenon, we expect that all of them are the result of the same cause. And if we also observe the same antecedent circumstances in all these instances, then we have reason to believe that this may be the cause we are seeking. The **METHOD OF AGREEMENT** helps us identify significant correlations between the phenomenon and antecedent circumstances, which then provide probable grounds to infer a causal relationship.

Suppose there is a dinner party attended by 10 people, and that six of the diners subsequently develop food poisoning. Naturally, we expect one of the food items served at the dinner party to be the cause. We can determine which item on the menu was responsible for the food poisoning by charting which items were eaten by each person, and then check to see if anything was eaten by the six people who became ill and not by any of the other guests. In this case, food poisoning is the phenomenon, all the items on the menu are the antecedent circumstances, and the 10 guests provide evidence of the cause. The chart might look like this (only "yes" answers are indicated in the appropriate spaces, "no" answers are left blank):

instances	salmon mousse	liver paté	garden salad	asparagus	cake	phenomenon (food poisoning)
Terry	yes	yes	yes	yes	yes	yes
Peter		yes	yes	yes	yes	
Rowan			yes	yes	yes	
Gillian	yes			yes	yes	yes
Polly		yes		yes	yes	
Eric	yes	yes		yes	yes	yes
Prunella		yes	yes		yes	
John	yes	yes		yes	yes	yes
Graham	yes		yes		yes	yes
Michael	yes		yes	yes	yes	yes

First, compare the two columns on the right. All those who developed food poisoning ate cake. But four people ate cake without getting food poisoning, which makes it unlikely that the cake is the cause. Next, compare the list of people with food poisoning and the list of people who ate asparagus. Not only are there instances of people who ate asparagus without getting sick, but one person who got sick (Graham) didn't eat that item. Again, we have reasons to eliminate this antecedent circumstance as the cause. For both the cake and the asparagus, there is no correlation between the list of people who ate it and the list of people who became ill. There is also no correlation between food poisoning and two other items on the menu. In fact, there is only one item that correlates with the people who developed food poisoning. All those who ate the salmon mousse became ill; furthermore, none of those who didn't eat salmon mousse became ill. This is a **SIGNIFICANT CORRELATION** because it is the only antecedent circumstance that correlates with the phenomenon. Therefore, there are good probable grounds to infer that the cause of the food poisoning was the salmon mousse.

The argument form associated with the Method of Agreement is as follows:

P occurs in 1 in circumstances x, y, z.
P occurs in 2 in circumstances x, z.
P occurs in 3 in circumstances x, y.
Therefore, it is probable that x is the cause of P.

In this schema, *P* is the phenomenon for which we are attempting to determine the cause. *1, 2,* and *3* refer to the instances of the phenomenon, and *x, y,* and *z* are the antecedent conditions. It says that when we look at all the observed instances and find only one antecedent circumstance that correlates with the phenomenon, then it is probable that the correlated circumstance is the cause.

Two qualifications about this argument form must be noted, however. First, the form isn't restricted to exactly three antecedent conditions or three instances of the phenomenon; it can be expanded to include more than three antecedent conditions or more than three instances. More importantly, if more than one antecedent circumstance corresponds with all instances of the phenomenon, as both cake and salmon mousse do in the food poisoning example, then the immediate conclusion to be drawn is *Therefore, it is probable that either x or y is the cause of P.* If only one of these antecedent circumstances correlates with the phenomenon, as the salmon mousse does, we should add the premise *But only x is correlated with P*, and then draw the conclusion as it is stated above. In fact, this is what we did when we ruled out cake as a possible cause of food poisoning.

As with other kinds of reasoning we have examined, it is important to be aware of the limitations of the Method of Agreement. In order to assess an argument based on the Method of Agreement we must check to see whether it exhibits any of the weaknesses to which arguments of this sort are vulnerable. First, it is possible that there will be agreement between the list of instances in which the phenomenon is observed and *two or more* antecedent circumstances. In these cases, we do not have a significant correlation; that is, we cannot narrow the list of potential causal factors down to one uniquely best candidate. It is also possible that two or more antecedent circumstances are *jointly necessary and sufficient* causes of the phenomenon. In these cases, we need to investigate further to identify a significant correlation before we can draw an inference. Second, it is possible that some *unobserved* factor may be playing a role in the instances being considered. An observed correlation provides only indirect evidence about the causal factor that is responsible for the phenomenon in question. Inevitably with this method, the antecedent circumstances that we select to observe will be determined by what we already think are potential causes. But it is possible that the real cause is something we simply do not expect and are not looking for. Third, even though we have a significant correlation in the dinner party example, we have not identified the causal factor precisely—i.e., we have not identified what is sometimes called the causal *agent*. We would want to know what kind of food poisoning was caused in this case (botulism, salmonella, staphylococcus, etc.), which ingredient in the salmon mousse harbored it, and how it got there. Answers to these questions will help explain why poisoning occurred in this particular dish (because ordinarily salmon mousse doesn't have this effect).

11.2.2 Method of Difference

When all the antecedent circumstances in two instances are the same except one, and a specific phenomenon is observed in one instance but not the other, then we have reason to believe that there is something significant about the single difference in the circumstances. The **METHOD OF DIFFERENCE** helps us identify the causal factor in one observed instance of a phenomenon by comparing it with a nearly similar instance in which the phenomenon is not observed. If we compare Terry and Peter at the dinner party, there is a significant difference between the two. Terry sampled from every dish being served, whereas Peter ate everything except the salmon mousse. So the *only* difference between what Terry ate and what Peter ate is the salmon mousse. The fact that Terry subsequently developed food poisoning and that Peter did not provides reason to believe that the mousse, the only item in Terry's meal that Peter did not also eat, was the cause of his food poisoning. Therefore, it is probable that the salmon mousse caused Terry's food poisoning.

The argument form associated with the Method of Difference is as follows:

P is observed in 1 in circumstances a, b, c, ... z.
P is not observed in 2 in circumstances b, c, ... z.
Therefore, it is probable that a is the cause of P.

In this schema, *1* and *2* refer to the two comparable things being observed. Again, *P* is the phenomenon for which we are seeking the cause, and the list of antecedent circumstances is *a* ... *z*. The argument says that because all antecedent circumstances for *1* and *2* are the same except *a*, we may infer that *a* is the cause of *P* in *1*.

There is a variation of the argument form associated with the Method of Difference that we must also consider. Often when the Method of Difference is used in experimental contexts, it is not particular cases that are compared; rather, *groups* of cases are compared. In these cases, a researcher begins with a uniform population, and then introduces a change into some members of the population while leaving the other members unchanged. Any subsequent difference between the two groups is inferred to be due to the change introduced to the one group. The change being introduced is called the **EXPERIMENTAL VARIABLE**, and the subsequent change to be observed is the **OBSERVED VARIABLE OF INTEREST**. The two groups are called the subject group and the control group.

The **SUBJECT GROUP** is the part of the population in which the experimental variable is being manipulated. The **CONTROL GROUP** is that part of the population in which the experimental variable remains unchanged. If there is a subsequent difference in the observed variable of interest between the two groups, then it is probable that the experimental variable played a causal role in producing that difference.

Suppose a biologist is able to grow a thriving population of a micro-organism species in a laboratory Petri dish. Suppose also that a divider is inserted into the center of the dish so that the two sides are now completely sealed off from each other, with the population on each side of the dish being the same size. We now have two comparable groups. One group can serve as the control by having all circumstances remaining constant. The other group can serve as the subject group, in which one antecedent condition is manipulated as the experimental variable. Consider what might happen to these groups when salt is added to the subject group. In this case, salt is the experimental variable, and population size within each group is the observed variable of interest. If the population sizes become noticeably different, then we may infer that the salt *caused* this difference. That is, we may infer that the presence of salt is responsible if the value of the observed variable of interest in the subject group is different from that of the control group. Moreover, we may make an inference of this kind as long as there is *any* significant difference between the two populations. If the size of the subject group diminishes, while that of the control group grows or remains constant, we may infer that salt caused the population reduction. Alternatively, if the population of the subject group is observed to grow while the population of the control group remains the same or diminishes, we may infer that salt caused the population growth. Finally, if no difference is observed after salt is introduced to the subject group, then we may infer that salt has no causal role to play in the life of this species.

The form of argument associated with the Method of Difference when used to compare groups in this way is much like the argument form described above in connection with particular cases.

> P is observed in the subject group in circumstances a, b, c, ... z.
> P is not observed in the control group in circumstances b, c, ... z.
> Therefore, it is probable that a is the cause of P.

As with other inductive forms of reasoning, the strength of these arguments is never purely a matter of form. Both premises could be true, yet the conclusion could be false. For the argument to be strong, the subject and control groups need to be designed in the right way. They must be comparable with each other, which is difficult to achieve when the population from which the groups are drawn is not as homogenous as the micro-organism population described above. If, for example, the population consists of human beings, then some members will be older than others, some taller than others, some healthier, some heavier, some male, some female, etc. How do we get comparable subject and control groups when the group members are so diverse? To begin, rather than aim to make every member of the two groups identical, we can aim to make the distribution of traits the same in both groups. We can assign to each group comparable numbers of tall people, healthy people, males, females, etc. This can be done in one of two ways: (1) the proportion of group members with a certain trait can be matched to the proportion of members in the other group with the same trait; or (2) the selection process can be randomized. **MATCHING** means ensuring that the percentage of females in the subject group is the same as the percentage of females in the control group, and so on for other traits. **RANDOMIZING** means assigning members to the subject and control groups without any regard for specific traits; if this is done arbitrarily, both groups should roughly equally reflect the distribution of traits within the overall population.

In order to assess an argument based on the Method of Difference, we must check to see whether it exhibits either of two weaknesses. First, we must consider whether there are any unobserved factors that may be playing a causal role in the occurrence of the phenomenon. As in the case of arguments based on the Method of Agreement, the antecedent circumstances that we notice will depend on what we think is relevant before we make our observations. But it is always possible that there might be other, unobserved differences that are causally more important than the ones we notice. Suppose, for example, that by mistake the salt added to the subject group in the Petri dish was iodized table salt rather than pure sodium chloride. Perhaps the observed change in the subject group was due to the iodine rather than the salt? Second, even if we have not missed anything relevant, we might not have isolated the specific causal factor involved, i.e., the causal agent. As we noted in connection with the Method of Agreement, there must be something in the salmon mousse that is the more precise cause of the food poisoning (botulism, salmonella, staphylococcus, etc.). These potential weaknesses, like the potential weaknesses with the Method of Agreement, are not reasons to distrust the method; rather, these are the reasons why we must take the conclusions of arguments based on these methods as being only probably true and not deductively certain.

11.2.3 Joint Method of Agreement and Difference

It is possible to use the Methods of Agreement and Difference together in one extended line of argument. In fact, with the **JOINT METHOD OF AGREE-**

MENT AND DIFFERENCE the two methods can yield mutually confirming conclusions. We can see how they complement each other in the dinner party example. The entire argument can be developed in two stages. First, we can consider all the people who got sick, using the Method of Agreement to uncover the correlation between the instances of diners who ate salmon mousse and the instances of those who developed food poisoning. If we look at only the people who got sick, however, two antecedent circumstances seem to correlate with the phenomenon. At this point, the Method of Difference can be used to determine if either of these potential correlations constitutes an actual correlation or a significant correlation. Terry and Peter are then compared using the Method of Difference.

A schematic account of how the two methods can be used together in the food poisoning example looks like this:

Stage 1: Method of Agreement

Food poisoning [P] *is observed in Terry* [1] *subsequent to eating salmon mousse* [a], *paté* [b], *garden salad* [c], *asparagus* [d], *and cake* [e].

Food poisoning [P] *is observed in Gillian* [2] *subsequent to eating salmon mousse* [a], *asparagus* [d], *and cake* [e].

Food poisoning [P] *is observed in Eric* [3] *subsequent to eating salmon mousse* [a], *paté* [b], *asparagus* [d], *and cake* [e].

Food poisoning [P] *is observed in John* [4] *subsequent to eating salmon mousse* [a], *paté* [b], *asparagus* [d], *and cake* [e].

Food poisoning [P] *is observed in Graham* [5] *subsequent to eating salmon mousse* [a], *garden salad* [c], *and cake* [e].

Food poisoning [P] *is observed in Michael* [6] *subsequent to eating salmon mousse* [a], *garden salad* [c], *asparagus* [d], *and cake* [e].

Therefore, it is probable that either the salmon mousse [a] *or the cake* [e] *is the cause of the food poisoning* [P].

Stage 2: Method of Difference (note that [2] at this stage is now Peter, not Gillian):

> *Food poisoning* [P] *is observed in Terry* [1] *subsequent to eating salmon mousse* [a], *paté* [b], *garden salad* [c], *asparagus* [d], *and cake* [e].
>
> *Food poisoning* [P] *is not observed in Peter* [2] *subsequent to eating paté* [b], *garden salad* [c], *asparagus* [d], *and cake* [e].
>
> *Therefore, by the Method of Difference it is probable that salmon mousse* [a] *is the cause of food poisoning* [P].
>
> *Therefore, by the Joint Method of Agreement and Difference it is probable that salmon mousse* [a] *is the sole cause of food poisoning* [P].

Because both arguments identify salmon mousse as the cause, after accounting for the data in different ways, the probability that we have concluded correctly in this case is greatly increased.

Now consider an example that uses the second version of the Method of Difference. In developmental genetics, experimental research can be conducted on genes by removing or altering a gene or part of a gene that is thought to be responsible for the development of a particular feature of an organism. If members of the subject group have genetic material "knocked out" and they subsequently develop in a way that is different from what is normal, then the missing genetic material is inferred to be responsible for normal development. In this research, the experimental variable is the genetic material, which is not changed in the control group but removed from members of the subject group. Suppose a particular gene sequence is knocked out in a subject group of mice and that all these mice subsequently grow abnormally short, dense bones. If the bones in mice of a control group grow normally, then the conclusion may be drawn that the missing genetic material is a *necessary* causal factor in the development and elongation of bones. Like the earlier example, the argument associated with this experiment can be analyzed as developing in two stages.

Stage 1: Method of Difference

> *Abnormally short, dense bones* [P] *are observed in the subject group of mice when they are deficient only in specific genetic material* [antecedent circumstances b, c, ... z].

Abnormally short, dense bones [P] *are not observed in the control group of mice who have the specific genetic material* [antecedent circumstances a, b, c, ... z].

Therefore, by the Method of Difference it is probable that the genetic material [a] *is the cause of bone growth.*

Stage 2: Method of Agreement

Short, dense bones [P] *are observed in mouse 1, which had gene X manipulated* [a], *was weaned at 3 weeks* [b], *fed a diet of sunflower seeds* [c], *and received regular exercise* [d].

Short, dense bones [P] *are observed in mouse 2, which had gene X manipulated* [a], *was weaned at 4 weeks* [e], *fed a diet of sunflower seeds* [c], *and received regular exercise* [d].

Short, dense bones [P] *are observed in mouse 3, which had gene X manipulated* [a], *was weaned at 3 weeks* [b], *fed a diet of corn* [f], *and received no exercise* [g].

Short, dense bones [P] *are observed in mouse 4, which had gene X manipulated* [a], *was weaned at 5 weeks* [h], *fed a diet of corn* [f], *and received regular exercise* [d].

Short, dense bones [P] *are observed in mouse 5, which had gene X manipulated* [a], *was weaned at 3 weeks* [b], *fed a diet of peanuts* [i], *and received no exercise* [g].

Short, dense bones [P] *are observed in mouse 6, which had gene X manipulated* [a], *was weaned at 5 weeks* [h], *fed a diet of peanuts* [i], *and received regular exercise* [d].

Short, dense bones [P] *are observed in mouse 7, which had gene X manipulated* [a], *was weaned at 4 weeks* [e], *fed a diet of millet* [j], *and received regular exercise* [d].

Therefore, by the Method of Agreement it is probable that the manipulation of gene X [a] *is responsible for the short, dense bones* [P] *in this group of mice.*

Again, we see in this example how the two methods can be used jointly to arrive at a conclusion. Not only can they be used jointly in this way, but the same set of observations can be used in both stages. In Stage 1 the subject group is considered in relation to the control group. Then in Stage 2 the members of the subject group are considered in relation to each other. These arguments are assessed by checking each stage for the same weaknesses that are associated with the method used in that stage.

11.2.4 Method of Concomitant Variations

When we observe that variations in two phenomena coincide with each other, we have reason to believe that they are causally related. The **METHOD OF CONCOMITANT VARIATIONS** helps identify correlations between two distinct phenomena. This method is useful for diagnosing the source of changes over time or across different populations. It is especially useful when the variations may be described in terms of a proportion. For example, as the temperature increases in a closed container of liquid, the internal pressure increases, while as the temperature decreases, the internal pressure decreases. Also, as we move to higher altitudes, the air pressure decreases, while as we move to lower altitudes, it increases. And finally, we might discover that the average number of cavities in children in various populations decreases when the amount of fluoride in the water supply increases. In the first case, the observed variations are *proportional*, which means that as one value increases, the other increases in a fixed proportion. In the second case, the observed variations are *inversely proportional*, which means that as one value increases, the other decreases. The third case is inversely proportional, too. Not all uses of this method are so easily described as mathematical proportions, however. Mill's own example of this method concerns the relationship between the ocean tides and the moon. In this case, there is a correlation between the position of the moon relative to the earth and the locations of high and low tides on the earth. On the basis of this observed correlation (along with other things we know about gravity), we infer that tides are caused by the moon.

There are two argument forms associated with the Method of Concomitant Variations. The form for proportional relations is as follows:

P1 occurs with P2 in 1.
P1+ occurs with P2+ in 2.
P1- occurs with P2- in 3.
Therefore, it is probable that changes in P2 cause proportional changes in P1.

In this schema, *1*, *2*, and *3* refer to the instances. *P1* is the phenomenon for which we are seeking the cause, and *P2* is the associated phenomenon that correlates with *P1*. It says that when the value of *P1* increases, so does the value of *P2* increase, and that when the value of *P1* decreases, so does the value of *P2* decrease. This form of the argument describes what happens when the method uncovers a proportional correlation. When the method uncovers an inversely proportional correlation, the form of argument associated with it is slightly different:

P1 occurs with P2 in 1.
P1- occurs with P2+ in 2.
P1+ occurs with P2- in 3.
Therefore, it is probable that changes in P2 cause inversely proportional changes in P1.

In this schema, the argument says that when the value of *P1* increases, the value of *P2* decreases, and when the value of *P1* decreases, the value of *P2* increases. From this it is inferred that the second phenomenon is inversely related to the first.

With this method, especially, a conclusion about the precise causal relationship between the two phenomena is difficult to establish. For while the Method of Concomitant Variations can identify a pattern of simultaneous or successive changes in the phenomena, this establishes only a correlation. We must go beyond the information in the premises to infer what is responsible for this correlation. For example, on the basis of available data about the tides and position of the moon alone, we cannot discount two other possibilities. It may be that the tides are responsible for the position of the moon, or it may be that a third factor is the cause of both. The first alternative simply reverses the causal order of the conclusion stated above. The second alternative says that Mill's sample argument about the tides commits the fallacy of common cause. We eliminate these possibilities, however, not because they are logically less secure than the original conclusion. Rather, because of what we believe about universal gravitation, we interpret tidal activity as the causal effect of the moon when we notice a correlation. Again, we are reminded that Mill's methods are inductive and that, however systematically they organize the data, they do not generate deductively certain conclusions. In order to function, this kind of reasoning requires a context within a larger body of knowledge about how the world operates. In this case, universal gravitation happens to be part of this body of knowledge.

11.2.5 Method of Residue

When we have a complex phenomenon the cause of which is partly explained by one or more antecedent circumstances, then we have reason to believe that any unexplained aspect of the phenomenon is caused by the remaining antecedent circumstances. The **METHOD OF RESIDUE** helps to identify the final causal factor in cases when the list of causal factors is nearly complete. For example, if we want to determine the mass of a certain quantity of liquid, we can weigh an empty container first, then weigh the container again with the quantity of liquid in it. Suppose the mass of the empty container is 1 kilogram, and that the mass of the container filled with liquid is 10 kilograms. By subtracting the mass of the container from the total mass of the container and liquid, we can determine that the mass of the liquid alone is 9 kilograms. More complex examples are possible. Suppose we want to determine what percentage of cases of hepatitis C have been transmitted sexually, and we have evidence to believe that there are three routes of transmission. If we know that 60 per cent of cases are caused by intravenous drug use, and that 15 per cent are caused by blood transfusions, then it is probable that the remaining cases (25 per cent) are transmitted sexually.

The form of the argument associated with the Method of Residue is as follows:

> P consists of parts P_1, P_2, and P_3.
> P occurs in circumstances a, b, and c.
> a causes P_1.
> b causes P_2.
> Therefore, it is probable that c causes P_3.

In this schema, P is a complex phenomenon with parts P_1, P_2, and P_3, while *a*, *b*, and *c* are the antecedent circumstances. It says that the phenomenon consists of three parts, with all but one part being caused by all but one of the antecedent circumstances. From this we can infer that the remaining part is probably caused by the remaining antecedent circumstance.

As with the Method of Difference and the Method of Concomitant Variations, the Method of Residue has a variant with a slightly different form. The second version of the method is used to identify an interfering factor when we are dealing with a well understood cause-and-effect relationship. Suppose a medication is ordinarily effective for reducing high blood pressure. If it is observed to be ineffective when taken at the same time as another medication, then the second medication is an interfering causal factor. The form of the argument associated with this version of the Method of Residue is as follows:

> Ordinarily, a causes P.
> P does not occur in circumstances a and b.
> Therefore, it is probable that b interferes in the causal relation between a and P.

In this schema, *P* is the phenomenon and *a* and *b* are antecedent circumstances. The argument says that while *a* usually causes *P*, it does not do so in the presence of *b*. On the basis of this, it is inferred that *b* probably interferes in the causal connection between *a* and *P*.

To assess arguments based on the Method of Residue, we must first determine whether the background causal relationship asserted in the first premise is sufficiently well understood to support the conclusion. While the other methods described by Mill infer a causal relation from a correlation, the Method of Residue infers something about *part* of a causal relation from an account of the *whole* causal relation. Of course, this concern about the first premise is not a *logical* problem, since it amounts to a concern about premise acceptability. Still, because this legitimate concern can never be eliminated, the inference to the conclusion using this method is only probable, it's not deductively certain. Secondly, not only does it remain possible that we do not understand the causal process as well as the first premise alleges, but it is also possible that the part of the phenomenon to be accounted for is entirely due to unobserved circumstances or to an unobserved circumstance *and* the residual observed circumstance. Perhaps the remaining 25 per cent of cases of hepatitis C infections are due to sexual transmission *and* some other source that has not been noticed before.

11.3 SELF-TEST NO. 17

For each of the following passages, identify which of Mill's methods is being used and describe the argument in the passage using the schematic form associated with that method.

1. Epidemiologists traced the illness [Severe Acute Respiratory Syndrome] back to a professor from China who was staying at Hong Kong's Metropole Hotel. Five other people who have come down with SARS stayed at the same hotel, with some of them staying on the same floor as the professor. (*CBC News Online*, In Depth: SARS Timeline [accessed on January 23, 2015])
2. The mortality rate of SARS in the general population is not as high as news reports might suggest, and usually it can be treated successfully. The most recent estimates peg the general mortality rate at 10–15 per cent. However, this is deceptive, for the mortality rate is not the same for all groups. For those who are otherwise healthy and under 25 years of age, it is slightly less than 1 per cent. But for the elderly and for those who are already suffering from respiratory conditions, the rate rises to

almost 50 per cent. So it is probable that SARS is much more difficult to treat successfully when the patient is elderly or suffering from a prior respiratory condition.

3. If the Fed [the Federal Reserve] raises the discount rate [at which it lends money to banks], banks cannot afford to borrow as heavily as before and have to curtail their lending and raise their own interest rates [to their own customers]. That results in less money flowing into the economy. Conversely, if the Fed relaxes its discount rate, financial institutions have more dollars for their customers. Seen from this perspective, the discount rate has a snowball effect: Raising it means that other interest rates go up as well and, other things being equal, economic activity slows down; lowering it has the opposite effect. (H.T. Reynolds, "The Federal Reserve System," http://www.udel.edu/htr/American/Texts/fed.html [accessed on January 23, 2015])

4. All domesticated dog breeds become quite adept at responding to human gestures such as pointing. By contrast, neither their genetically close relatives, wolves, nor intellectually superior primates, such as chimpanzees, respond as well as dogs to the same cues. Therefore, it is likely that something happened in the evolutionary heritage of dogs (i.e., a combination of breeding and socialization) since they split off from wolves and that factor is responsible for this ability.

5. Former US Olympic hockey coach Herb Brooks probably fell asleep at the wheel before his fatal car crash last month, according to a report released yesterday [by] the [Minnesota] state police. The report confirmed that Brooks wasn't drinking, speeding, talking on his cell phone, or having a health problem before the crash. Police said weather and road conditions were ruled out [as the cause]. (*Associated Press*, "Brooks Most Likely Asleep Before Crash," Sept. 17, 2003)

6. Twins have a reduced risk of suicide, which supports the hypothesis that strong family ties reduce the risk for suicidal behavior. This finding was consistent across cohorts, sex, and zygosity. As we used population based register data there was little room for selection bias. The strongest risk factor for suicide is mental illness, but other Danish register studies have found mental illness to be slightly more common among twins than among singletons. This should lead to a higher proportion of twins committing suicide compared with the general population, but our findings show exactly the opposite, further underscoring the

importance of strong family ties. (Tomassini, Cecilia et al., "Risk of Suicide in Twins: 51 Year Follow up Study." *BMJ: British Medical Journal* 327.7411 (2003): 373–74)

7. Stressed children are more likely to go for high-fat foods and snacks than their placid peers, regardless of whether they respond to anxiety by eating more or less than usual. A study of 4,320 British schoolchildren, appearing in the journal *Health Psychology*, found a strong relationship between stress and fatty foods. Those 11-year-olds who were the most stressed ate nearly twice as much fatty foods as their less anxious classmates. They also were the biggest snackers. At the same time, they "were also less likely to consume the recommended five or more fruits and vegetables a day and eat a daily breakfast," said Jane Wardle, director of the Cancer Research UK Health Behaviour unit. (*Globe and Mail*, "Social Studies: Stress and the Munchies," A22, September 3, 2003)

8. In a famous experiment conducted … by the Stanford University psychologist Philip Zimbardo, a car was parked on a street in Palo Alto, where it sat untouched for a week. At the same time, Zimbardo had an identical car parked in a roughly comparable neighborhood in the Bronx, only in this case the license plates were removed and the hood was propped open. Within a day, it was stripped. Then, in a final twist, Zimbardo smashed one of the Palo Alto car's windows with a sledgehammer. Within a few hours, that car, too, was destroyed. Zimbardo's point was that disorder invites even more disorder—that a small deviation from the norm can set into motion a cascade of vandalism and criminality. (Malcolm Gladwell, "The Tipping Point" in *The New Yorker*, June 3, 1996)

11.4 INFERENCE TO THE BEST EXPLANATION

In section 10.4 we examined how Induction by Confirmation can be used to explain a set of observations. At the same time, observations can be used to test the truth of a hypothesis. When an observation statement that has been deduced from a hypothesis turns out to be true, we have inductively confirmed the hypothesis. When the observation statement is not true, we have disconfirmed the hypothesis. As we saw in Chapter 10, inductive confirmation falls short of deductive certainty, and a disconfirmed hypothesis can be saved or revised in a number of ways. So while we are concerned with the truth of a hypothesis, we do not have a direct and conclusive test to establish whether it is true or false. Moreover, a good hypothesis is not only true, it is also explanatorily adequate. **EXPLANATORY ADEQUACY** means that the hypothesis explains *all* instances of the phenomenon, and that it does so with as much precision as possible. The first component of explanatory adequacy concerns the **SCOPE** of the hypothesis; that is, it tells us how many instances of the phenomenon the hypothesis explains and specifies what restrictions apply. Naturally, the wider the scope the better the hypothesis. The best hypotheses have a universal scope; in other words, they cover *all* instances. The second component concerns the **ACCURACY** of the hypothesis; that is, it tells us how much detail the hypothesis is able to provide as an explanation of any particular observation. Naturally, the more detailed a hypothesis the better, for it explains each instance precisely. For a hypothesis to be completely explanatorily adequate, there should be no unexplained instances of the phenomenon (in which case, we say its scope is *universal* or *unrestricted*), and the observation statements should be deducible from the hypothesis with as much precision as we are able to achieve in making the observations themselves.

Of course, if every observation is explained by a single hypothesis, then we have good grounds to accept the hypothesis. But there are two common situations in which the evidence is not so decisive. (1) Sometimes two or more **RIVAL HYPOTHESES** may each purport to explain *some* observations without either one explaining all of them; that is, both hypotheses are restricted in scope. If the hypotheses are inconsistent with each other, then at least one of them is not true. How do we decide which one to accept? (2) Sometimes two or more rival hypotheses purport to explain *all* the observations; that is, both hypotheses are unrestricted. Again, if they are inconsistent with each other, then at least one of them is not true. How do we choose which of these to accept?

The most decisive strategy for resolving an impasse of either kind is to devise a crucial experiment. A **CRUCIAL EXPERIMENT** is designed to generate an observation that is predicted to occur by one of the rival hypotheses and predicted not to occur by the other(s). In this way, one rival is confirmed and the other(s) disconfirmed. However, a disconfirmed hypothesis may be qualified, some of its key concepts may be redefined, its scope may be limited, or another part of the theory that generated it may be revised (see section 10.4). For these reasons, it is difficult to eliminate any of the rival hypotheses once and for all. There are, after all, people who still maintain that the world is flat and not spherical. Still, this doesn't mean that all hypotheses are equally acceptable or that there are never reasons to favor one hypothesis over others. Usually, we must assess hypotheses by asking which, on balance, is better than its rivals. The components of explanatory adequacy provide standards for making such comparative assessments. We can determine which of the rivals has the widest scope, and we can determine which is more accurate.

In section 11.4.1 we will consider what rational grounds are available to decide which hypothesis to prefer when two restricted hypotheses compete for our acceptance. Then in section 11.4.2 we will consider what rational grounds are available to decide which to prefer when two unrestricted hypotheses compete to explain the observations. These are not purely theoretical disputes, for they go right to the heart of what a science is supposed to do. Sciences are supposed to provide explanations for observed phenomena, and we want to use the best explanatory hypothesis available when we formulate an explanation. So it is important that we find a way to settle these disputes in a rationally defensible manner.

11.4.1 Choosing between Rival Restricted Hypotheses

Rarely, if ever, is there only one hypothesis that purports to explain something. When we have more than one hypothesis, we might say that the different hypotheses compete with each other for our acceptance. In these cases, we need to choose between rival hypotheses. Suppose Louis notices that the last piece of chocolate birthday cake that he has been saving in the fridge is gone and that the plate it was on is sitting in the sink, licked clean. If the only people who have been in the house since he put away the cake are his two brothers, Leo and Paul, then Louis might formulate two rival hypotheses. These are as follows:

H1: *Leo ate the last piece of cake.*
H2: *Paul ate the last piece of cake.*

Suppose also that Paul cannot stand the taste of chocolate or cake, and that chocolate cake is Leo's favorite dessert. Bearing this in mind, Louis will be able to recall many occasions when Leo devoured chocolate cake enthusiastically, and countless times when Paul declined cake (and other chocolate treats). All these memories of Leo eating cake can be counted as confirming observations for H1, and the memories of Paul declining can be counted as disconfirming observations of H2. In this case, the choice between H1 and H2 is pretty easy. H1 is far more likely to be true than H2. It isn't always as easy to choose between rival hypotheses as it was for Louis. But, in general, the logical situation is not fundamentally different from this. Let us develop a more sophisticated example to illustrate how this is so.

In the seventeenth century, physicists were divided over a controversy about the nature of light. Isaac Newton maintained that observations about the movement of light were best explained by thinking of it as an emission of particles. Christian Huygens, on the other hand, maintained that observations were best explained by thinking of its motion as a traveling wave. A summary of this controversy, which lasted until the early years of the twentieth century, illustrates the problem that arises when restricted hypotheses compete with each other (i.e., when some observations are explained by each hypothesis, but neither hypothesis explains all of them).

Both camps appealed to observations that seemed to confirm one hypothesis but not the other. According to Newton and other particle theorists, the hypothesis that light moves as a particle is confirmed by the observations that light travels in a straight line and that it reflects off mirrors in the same way a ball bounces off a wall. And according to Huygens and other wave theorists, the hypothesis that light moves as a wave is confirmed by observations that two beams of light can pass through each other without either one being disturbed and that light passing through a double-slitted screen diffracts in a pattern that resembles waves on a lake passing through two openings in a barrier. At the same time, however, the scope of both hypotheses is limited. The particle hypothesis cannot explain diffraction, and the wave hypothesis cannot explain why light does not move around corners.

The two hypotheses and their respective confirming observations can be described using the argument form for Induction by Confirmation as follows:

H1: Light as a Particle

If light is a particle [h1], *then a beam of light will travel in a straight line* [o1], *and a beam of light striking a mirrored surface will reflect off the surface at the same angle as the angle of incidence* [o2].

A beam of light does travel in a straight line [o1].

A beam of light striking a mirrored surface does reflect at the same angle as the angle of incidence [o2].

It is probable, therefore, that the hypothesis that light is a particle is true [h1].

H2: Light as a Wave

If light is a wave [h2], *then two beams of light will cross paths without disturbing each other* [o3], *and a beam of light passing through a double-slitted screen will diffract* [o4].

Two beams of light do cross paths without disturbing each other [o3].

A beam of light passing through a double-slitted screen does diffract [o4].

It is probable, therefore, that the hypothesis that light is a wave is true [h2].

If one of these hypotheses had been more accurate than the other, then that might have tipped the scales in favor of the more accurate hypothesis. But in this case, it was not possible to settle the dispute in this way. Consequently, because neither hypothesis explained all the observations, the challenge for physicists was to find a formulation of one hypothesis that would expand its scope to cover the observations previously associated with the other hypothesis. The hypothesis with the greater scope will then have more explanatory power, which tips the scales in its favor.

Essentially, the wave hypothesis gained wider acceptance over the course of the eighteenth and nineteenth centuries because it was able to explain *o1* and *o2*, while the particle hypothesis could not accommodate *o3* and *o4*. However, waves require a medium, just as water is the medium of waves on a lake, and an auxiliary hypothesis about the nature of this medium was accepted along with the wave hypothesis; it is, in short, a presupposition of the wave hypothesis. An **AUXILIARY HYPOTHESIS** is a second hypothesis that is logically required by the hypothesis under investigation. In this case, the auxiliary hypothesis postulated that all space (even a vacuum through which light can travel) is permeated with ether, a transparent, intangible, massless substance that acts as a stationary medium for the propagation of light waves. It was this auxiliary hypothesis that eventually raised concerns about the wave hypothesis. In 1895 two American scientists, Albert Michelson and Edward Morley, performed an experiment that disconfirmed the existence of ether. Huygens's formulation of the wave hypothesis could not be sustained without the auxiliary hypothesis about ether (although, as it turns out, the real change concerned the assumption that waves require a medium in which to travel). Because the particle hypothesis still could not accommodate *o3* and *o4*, it became necessary for physicists to look for a new hypothesis about the nature of light. Accordingly, the new hypothesis had to accommodate all the observations about the phenomena associated with light (which by this time was believed to be only the visible part of the range of the electromagnetic spectrum) and to overcome the duality of the wave/particle controversy. The new hypothesis, that light is a quantum of electromagnetic energy, became a central part of quantum theory.

For our purposes, there is no need to dwell on the technical, theoretical details of this controversy and the transition from the classical seventeenth-century hypotheses about light to the contemporary, quantum mechanical hypothesis. Our interest in this controversy lies in tracing out the rational principles involved in the process of searching for the best explanation. Most importantly, this case illustrates how a dispute between two rival restricted hypotheses can be conducted as a rational debate and how its resolution can be described in rational terms. We must assess the rival hypotheses using the components of explanatory adequacy. We can do this by asking two questions: (1) is one hypothesis more accurate or precise than others? If not, then (2) does one hypothesis cover more observations than others? And, of course, if one hypothesis is both more accurate and covers more observations, it is that much better.

11.4.2 Choosing between Rival Unrestricted Hypotheses

Since the middle of the nineteenth century, evolutionists and creationists have debated over how to explain the existence of fossils and certain other geological phenomena. On one side, evolutionists claim that fossils are the petrified remnants of plants and animals that died millions of years ago. A central hypothesis of evolutionary theory is that the earth is millions, indeed billions, of years old and that this explains the observed fossil record. On the other side, a popular hypothesis among creation theorists is that the earth is only approximately 6,000 years old, and fossils are explained as taking their present form in the original moment of creation. A summary of this controversy illustrates the problem that arises when two rival hypotheses compete to explain *all* the same observations. Because this debate is still ongoing, and because it would divert us from our purpose to enter the fray as participants, we won't attempt to trace it through to a resolution (as we did in section 11.4.1). Instead, we'll use the debate to explore one further dimension of scientific reasoning.

In the wave/particle debate the two camps were initially divided not only by their respective hypotheses, but also by the observations that each side deemed to be relevant for testing the hypotheses. In the evolution/creation debate the disagreement may be even more fundamental, for both sides appear to accept the same observations as relevant. The fundamental difference can be shown by considering the following descriptions of the two lines of reasoning using the form of Induction by Confirmation.

Evolution

If the earth is millions of years old [h1], *then part of the lithosphere (i.e., the rocky crust of the earth) will be stratified* [o1], *and there will be fossils within these layers* [o2].

Parts of the lithosphere are stratified [o1].

Fossils are found in the layers [o2].

Therefore, it is probable that the earth is millions of years old [h1].

Creation

If the earth was created 6,000 years ago [h2], *then part of the lithosphere will be stratified* [o1], *and there will be fossils within these layers* [o2].

Part of the lithosphere is stratified [o1].

Fossils are found in the layers [o2].

Therefore, it is probable that the earth was created 6,000 years ago [h2].

How could two incompatible hypotheses both be confirmed by the same observations in this way? This is one of the most difficult and controversial theoretical questions associated with scientific reasoning. Fortunately, it's not necessary for us to solve this problem here. For our purposes, it will be enough to outline what lies behind the division between the two hypotheses and to indicate how the debate over a controversy of this sort can be conducted.

Because *o1* and *o2* are worded as neutral descriptions of facts, it looks as if evolutionists and creationists are both making arbitrary deductions in the first premise of each argument. But, as we pointed out above, explanatory adequacy requires that the observations deduced from a hypothesis be both true for all observation statements and *precise*. It is in attempting to meet the second part of the standard of explanatory adequacy, precision or accuracy, that the deep difference between these two hypotheses emerges. Neither evolutionists nor creationists should be content with the bland, undetailed description of the facts as *o1* and *o2* describe them. For example, consider how each would elaborate on *o2* to improve its precision. Each camp will define *fossil* differently. An evolutionist might define it as *the petrified remains or traces of any organism that lived in the distant geological past*, whereas a creationist might define it as *mineral patterns that have the appearance of plants or animals*. Moreover, an evolutionist will maintain that petrification of organic material takes place under certain conditions in a process that requires millions of years to complete, whereas a creationist will maintain that the mineral patterns that distinguish fossils were produced by God in the act of creation. Now we begin to see how the observation statements can be explained so differently by the two hypotheses: what evolutionists and creationists disagree over is the *interpretation* of the observed phenomena. *This* is the deep problem that divides evolutionists and creationists so sharply.

In section 10.4 we saw that hypotheses can be disconfirmed, but also that a disconfirmed hypothesis can be saved in a number of ways—for example, by restricting its scope. Then in section 11.4.1 we saw that when two incompatible restricted hypotheses compete for acceptance, the balance of confirming and disconfirming instances can tip the scales in favor of one over the other. This was how the original particle hypothesis was eventually ruled out as an explanation of light by the late nineteenth century. We also saw in section 11.4.1 that a hypothesis can come to be rejected if a necessary auxiliary hypothesis is unacceptable. This is how the original wave hypothesis also lost favor among physicists. In all these cases, hypotheses were tested by determining whether the observation statements that could be deduced from them were true. Now we see that in order for two scientists to come to the same assessment of a hypothesis, each must interpret the observed phenomena in the same way. Differences of interpretation will become evident in the way observational details are described with precision. Moreover, the precise way these details are described depends on the entire theory within which the hypothesis is set.

A **THEORY** is a systematically integrated set of general principles, methods of investigation, and concepts whose function is to explain a wide array of phenomena. Theories generate hypotheses about specific phenomena, but, significantly, they also provide the filter or lenses through which we interpret the observations that test hypotheses. (This is quite different from the popular, unscientific conception of a theory as an untested hypothesis or a guess, by the way; in science, a theory is much more systematic and regulative than these popular notions suggest.) The interpretive role performed by a theory in formulating a precise observation statement is the ultimate source of some profound controversies in science, such as over the age of the earth. For the most part, the rivalry between the wave hypothesis and the particle hypothesis was a rivalry between two players in the same game, so to speak; both hypotheses attempted to explain the nature of light within the larger theoretical framework of a *mechanistic* theory of the universe. The mechanistic theory attempts to explain all phenomena as material elements that interact in a mathematically determined fashion. Because both waves and particles can be described in these terms quite well, both were attractive models for the nature of light. But in the debate over the age of the earth, the two rival hypotheses may not be playing the same game, which is illustrated by the disagreement over how to formulate observation statements precisely; rather, what may be ultimately at stake here concerns which game rules to apply. If two games are very similar (e.g., ice hockey with body-checking and ice hockey without body-checking), then it may not be difficult to debate over which rules apply. If two games are similar in some ways but quite different in others (e.g., ice hockey without body-checking and lacrosse), then this debate is more difficult. And if two games are very different (e.g., ice hockey with body-checking and golf), then the debate is going to be much more difficult, if not impossible.

When theories differ so much that it's impossible to test any of their respective hypotheses with the same observation statements, then we must consider the possibility that the theories are incommensurable. **INCOMMENSURABILITY** literally means *not capable of being measured by a common standard*, and it refers to the relationship between two things that are so different that it's not possible to make an objective comparative judgment about them. (In essence, this is what the cliché about the impossibility of comparing apples and oranges means.) Theories are incommensurable when there is no available standard that is independent of both theories and by which we can test to determine which theory is better. It may well be that evolutionism and creationism are indeed incommensurable. This itself is a matter of much debate. But within the scientific community, evolution has been the preferred theory, and along with it, the hypothesis that the earth is millions (indeed billions) of years old. How has this happened?

Evolutionism has simply offered more opportunities to pursue research into the questions that give science its distinctive purpose, that is, the *Why ...?* questions about the way natural processes work. No decisive refutation of creationism has been involved in this preference. And because theories can be saved from refutation in all the same ways that hypotheses can be saved from disconfirmation, no such refutation should be expected. Still, this doesn't mean that the preference for one theory over the other, such as the preference for evolution over creationism, is arbitrary. For in this case, the research program offered by creationism seems always to lead to the same answer. For creationism, all the *Why ...?* questions are ultimately resolved in terms of God's intentions, which are often said to be beyond human comprehension. It's not that there aren't interesting and thoughtful ways to provide answers to such questions with creation theory. But, however interesting and thoughtful they may be, they have not been interesting to scientists because the justification for them rests as much, if not more, on biblical exegesis than on empirical investigation. Creationism does not suggest observation statements that are testable by empirical scientific research. In contrast, evolutionism suggests many avenues of empirical research. For example, if one genus of animal is hypothesized to have evolved from another genus (e.g., birds may have evolved from reptiles), then we should find no evidence of the later evolving genus below a certain point in the layers of rock and we should find some evidence of the earlier evolving genus below that point. It is primarily because evolutionism has generated so many detailed hypotheses of this sort that it has been accepted by most scientists.

Here again, we see that the standard of explanatory adequacy provides guidance in the scientific choice between two hypotheses, even when the hypotheses are unrestricted and may be incommensurable. In these cases an assessment can be made using the standard of accuracy: the hypothesis that generates more precise observation statements to confirm it is more explanatorily powerful.

11.5 CASE FOR DISCUSSION: SEMMELWEIS'S DISCOVERY OF ANTISEPSIS

The process of discovery recounted below may be analyzed using several of the concepts and argument forms covered in this chapter, along with a few from earlier chapters. There are several discrete stages in the overall line of reasoning that begins with the first observation of the phenomenon and ends with a conclusion about the cause of that phenomenon. Using the information from this account, (1) identify the hypotheses being tested and (2) identify the separate stages in the overall line of argument. Also, (3) identify what kind of argument form is used at that stage and (4) describe the argument using the appropriate form for each stage.

Ignaz Semmelweis (1818–65) was a professor of medicine in the Doctor's Ward of the Vienna General Hospital from 1844 to 1848. The hospital was dedicated exclusively to the obstetrical care of poor women. For many years before Semmelweis arrived and for many of his first months there, the hospital had an unusually high mortality rate among the patients who came there to deliver their babies. In particular, there were frequent outbreaks of a life-threatening illness known as puerperal fever, which is better known as childbed fever.

We now know that childbed fever is a bacterial infection of the female genital tract that is contracted during childbirth. In the nineteenth century, however, the source and transmission of infectious disease was very poorly understood. It was not until 1857 that Louis Pasteur discovered germ theory, which explained both the specific cause of infectious diseases as being microorganisms and how these diseases are spread in a community. Without Pasteur's theory, it was much more difficult to understand childbed fever. (Even then, it was not until the 1870s that Pasteur himself identified the microorganism responsible.) Furthermore, in the 1840s there was no treatment for the condition, and very few women who contracted it survived.

The hospital had two divisions, the Doctor's Ward where Semmelweis and other professors delivered babies and taught obstetrics students, and the Midwives' Ward where midwives and their students delivered babies; the two divisions were in separate but adjacent wings. Besides providing obstetrical training on live births, physicians in the Doctor's Ward performed autopsies as part of the course of instruction for their students. For many years the Doctor's Ward had a maternal mortality rate of approximately 10 per cent, which means that of the 3,000 or so births in any year 300 women died. During epidemic periods the maternal mortality rate rose as high as 20–50 per cent. In the Midwives' Ward, however, the rate of death due to childbed fever was much lower. In 1844 it was 2.3 per cent, in 1845 it was 2.0 per cent, and in 1846 it was 2.7 per cent.

When Semmelweis raised concerns about the rate of childbed fever in the Doctor's Ward and asked his supervisor how it could be so high, he was told that the fever was inevitable with patients at this hospital, who were among the poorest in Vienna (expectant women from higher economic classes were generally attended to by a private physician in their own homes). He was told that the women treated in the Vienna General Hospital were from an "inferior" class of society, which made them susceptible to *miasma* (which means *pollution*), or atmospheric forces that harbored childbed fever in the air. Semmelweis found this explanation unsatisfactory for two reasons: (1) women who gave birth on the way to the hospital were far less likely to die from childbed fever than those who actually made it there in time to give birth (this rate was nearly 0 per cent). Why would the atmospheric conditions in the neighborhood outside the doors of the hospital be different from those inside the hospital? (2) Since every woman admitted to the hospital came from the same socioeconomic class and the two wards were adjacent, *all* the women at the hospital should have been equally affected by the *miasma*. But the infection rate in the Midwives' Ward was much lower than in the Doctor's Ward, despite the fact that both divisions were equally exposed to the "atmospheric conditions" that were said to be responsible. Why was there a significantly different rate of infection between the two divisions?

Someone suggested that overcrowding in the Doctor's Ward might be the problem. But the Midwives' Ward was more crowded than the Doctor's Ward, so this couldn't explain the difference. (In fact, many women contrived to be admitted to the Midwives' Ward, because the rate of maternal mortality in the Doctor's Ward was notoriously high.) Another suggestion was that foreign medical students in this division (of which there were many) were too rough with women during examination. But, Semmelweis noted, delivery was far more traumatic than anything that might be done during examination; besides, there was no evidence that foreign students were rougher than other students or the physicians themselves. Semmelweis also noted that there seemed to be no differences in the patients of the two divisions with respect to the diet or general care. Finally, Semmelweis noticed that there was a higher incidence of childbed fever in the Doctor's Ward among women who had long labors than those whose labors were brief.

Looking for what may account for the differences between the two divisions, Semmelweis noticed that priests coming to perform last rites in the Doctor's Ward entered at one end of the hall, and then proceeded down the entire hallway, past all the beds of new mothers to the room to which infected patients were removed after being diagnosed. By contrast, a priest entering the Midwives' Ward to perform last rites could get to the dying patient without passing any other patients. Perhaps the psychological stress of having the priest pass by otherwise healthy women in the Doctor's Ward to perform this task was provoking periodic epidemics? Semmelweis arranged for the priests to enter by a secondary staircase hoping that this would reduce the infection rate. No reduction in the rate of infection was observed. Other, similar changes were introduced to the protocols of the Doctor's Ward, all equally ineffective.

Early in 1847 one of Semmelweis's colleagues, Jakob Kolletschka, died after receiving a puncture wound from a scalpel while performing an autopsy. While examining the pathology report on Kolletschka's death, Semmelweis noticed that his colleague exhibited the same symptoms as childbed fever. Semmelweis thought that "cadaveric matter" had entered Kolletschka's blood through the scalpel wound and caused his illness. Kolletschka's death led Semmelweis to believe that the women under his care had died from a similar kind of poisoning, not from something in the air. Semmelweis, his colleagues, and their students had been inadvertent carriers of the infectious "cadaveric matter." They often came to deliver babies directly from performing autopsies, and routinely they examined patients after only superficially washing their hands. In contrast, the midwives never performed autopsies; therefore, they were not infecting their patients with "cadaveric matter."

Semmelweis ordered physicians and students to wash their hands in a chlorinated lime solution as a chemical means to disinfect their hands after performing autopsies. The results were immediate and dramatic. The maternal mortality rate had been 13.1 per cent in July 1846, and 18.05 per cent in August 1846. In May 1847, just before Semmelweis's antiseptic hand-washing practice was introduced, it was still as high as 12.24 per cent. But in the first month after the practice was instituted, June 1847, it dropped to 2.38 per cent. The rate dropped in July 1847 to 1.2 per cent, and in March and April 1848 there were no deaths due to childbed fever in the division. By the end of 1848 the annual mortality rate due to childbed fever in the Doctor's Ward fell to 1.27 per cent (correspondingly, it fell to 1.3 per cent in the Midwives' Ward).

For many months after the practice was introduced in the Doctor's Ward, the infection rate of childbed fever remained close to this low level. Then, after a period of stability, a wave of infections temporarily spread through the division. One day during rounds, physicians first treated a pregnant woman who had an ulcerated cervical tumor, and then they proceeded to examine several other women immediately afterwards. Since the physicians had already disinfected their hands before treating the first patient, they didn't disinfect them again afterwards. Eleven of the 12 women treated after this woman died of childbed fever. In light of this incident, Semmelweis revised his initial claim about the source of infection. Now he maintained that it was caused by "cadaveric matter or putrid organic matter." We now know that the more precise description of the cause is microscopic bacteria, but Semmelweis's imprecise description was enough for practical purposes. As long as physicians and students in the division washed their hands with the chlorinated lime solution before examining any patient, the rate of infection remained low.

Independently of Semmelweis's work, Oliver Wendell Holmes (1809–94, poet and father of the famous jurist, Oliver Wendell Holmes, Jr.) came to the same conclusions about both the contagiousness of childbed fever and how to prevent it. And shortly after Semmelweis's death, Joseph Lister (1827–1912) designed a similar antiseptic system for clinical practice based on Louis Pasteur's germ theory. Semmelweis is now acknowledged to have been at the leading edge of this revolution in hospital practice.