**Arguments to the Best Explanation.** If we are to make any further progress, then, we need to consider what the best explanation is for clearly observable order in the world. If it turns out that the best explanation is one developed from God's existence, then we will have an argument, of some strength at least, for God's existence. We will not have a deductive argument, however, since the conclusion of our argument could be false even if this is unlikely while the premises are true. Sometimes the best explanation that we have for some observation is later discovered to be false.

The argument from design was a fairly good argument at one time, because at that time design was the best *available* explanation for the functionality of living things. But with the development of the theory of evolution in the 19th century, the situation changed. Evolution explains the apparent adaptation of means to ends in biology. Evolution has now established itself as an extremely good explanation of biological function, and a far better one than the theistic hypothesis that an omniscient, omnipotent, perfectly good designer made animals and plants. To see all this more clearly, we need to examine arguments to the best explanation more closely.

An argument to the best explanation runs like this:

(1) Of the possible explanations we've been able to come up with for facts $F_1$ through $F_n$, the best is E.
$\therefore$ (probably) (2) E is true.

How probable will the conclusion be? This will depend on how well E explains the fact, and how much better it is than the other possible explanations. It might be quite good at explaining the facts, but have some very good competitors as well, or it might not be very good at all, even though it is the only explanation anyone can think of. In both these cases the probability it is true may be low. If it is the *only possible* explanation (given the background), it must be true, as Sherlock Holmes pointed out, no matter how problem-filled it is, but one needs to be sure it really is the only possible one.

The strength of the argument will also depend on the nature of the facts to be explained. If it explains a great many facts in considerable detail, an explanation is more likely to be true than if it explains only a small number of rather general facts.

If a theory explains the facts, each of the facts can be seen to follow (not necessarily deductively) from the assumption that the theory is true, together with various other things we know about the world, our background information. If a theory is the correct explanation, then the facts will be true *because* the theory is true, and given how the world is otherwise, they would not have been true had it not been. This is why we can regard the facts explained by the theory as clues to the true theory. The theory reports the cause or reason why the facts are true. So, if he really is the murderer, the clues are there because of the murder he did, not for some other reason, and if he had not committed the murder, the clues would not have been there.

It should be noted that each clue is explained by the theory independently, quite possibly with a different background than is relevant in the explanation of other clues. Moreover, a clue can support the theory without supporting every element of it, and different clues may support different elements. So there may be a clue that the butler was present in the room which would have been there even if the butler had not done the murder at that time. This becomes a significant clue, and is explained, within the context of doubt about where the butler might have been at the time of the murder. Given how the murder in fact occurred, the Butler had to be there,

and this is something which is best explained by his being there, and so supports, within one context of doubt, our theory that the Butler did it. The set of facts that we list to be explained in any particular argument to the best explanation, then, will depend on the history of the previous discussion, on what is or is not reasonably to be doubted. A full account of the evidence for any present day scientific theory will be an historical one, and the claim to be made for the theory will be that it has proven the best theory over the history of the discussion, by explaining better than its competitors at each stage, or the most important or most recent stages, of the discussion.

The best explanation available to us is not always the correct one. It might fail in a number of ways. The correct explanation might not be one we have formulated yet—perhaps we aren't even aware the father–in–law could be a suspect. Our background information might include false notions that lead us to think that a theory explains a fact when, given accurate background information, the theory would in fact predict its falsehood, or be irrelevant to it. The meaning of a clue, a fact to be explained, might change when we correct or expand our background information, what looked significant may now look insignificant, what implicated the maid may now implicate the butler—or we might discover that one of the supposed clues is not true at all when we find that the apparently credible witness lied. The hitherto best explanation might drop out of contention when new, relevant facts are uncovered—new clues may lead us to reassess the case completely.

To assess how good an explanation is, it is useful to look at the following requirements:

**It should explain the facts to be explained.** It may be that no explanation handles all the facts, and in that case one should tentatively opt for the one that explains the most and/or the most important facts, which would include those that bear on the most important doubts, and set off this theory most effectively against its competitors. What facts are most important will change as the investigation continues. An explanation may explain a fact better or worse: if the explanation depends on a non-deductive argument working from the explanation to the fact, the argument in question is stronger or weaker, and if it depends on a deductive argument, there may be premisses of that argument which are more or less probable.

**It does not predict that things are true that we can ascertain are false.** A theory may be forgiven false predictions if those predictions are of relatively unimportant facts, or based on fairly weak arguments with uncertain background assumptions, but a spectacular failure here may make it reasonable to reject a theory outright.

**It does not contradict what we accept already as true.** Even better, what we already accept as true in other areas may make the explanation a plausible one. Sometimes an explanation uses only those facts we already knew, introducing no new facts or theories, but rather pointing out the hitherto unnoticed relevance of facts and theories we already know. On other occasions an explanation will require us to give up strongly entrenched beliefs, introducing notions that are entirely new to us, perhaps in response to facts only recently uncovered. In the first case we have no trouble accepting the explanation, but in the second we will, reasonably, resist accepting it until other, more conservative explanations have been shown to be utterly inadequate. In the same vein, it may be that the explanation fits a pattern of explanation we have seen to work well in other areas, so there is sometimes cross–fertilization between fields of knowledge when a successful strategy of explanation in one area is applied in another. Sometimes we discover a causal pattern which is found in a number of different areas of research, and its presence in a specific area will seem more likely if we find it often elsewhere. One such causal pattern is the operation of dynamic equilibria. The presence of 'restoring forces'

and dynamic equilibria in physical systems such as pendulums makes it plausible to look for them in economic systems.

**It explains, or even better, predicts, new facts that it was not designed to explain,** and for which we have no other explanation ready to hand. So we may feel sure that we've broken a code when we can decode new messages not in the original group of messages that we studied to break it.

**A successful explanation may suggest new explanations in other areas**, perhaps following the same causal pattern, or perhaps because the explanation leads to the discovery of new facts which lead to solutions of other problems in a kind of chain reaction. The explanation is "theoretically fruitful."

**It is simple, but not too simple.** We need to generate predictions, at least predictions with some probability of being true, and so we are unable to test theories so complex that we cannot be sure what follows from them. To advance such theories may be tempting, but until we get in a position to see what follows from them more accurately, we may not be able to say how likely they are to be correct. On the other hand, the explanatory theory needs to be clearly formulated, detailed, so that it explains as much as possible in as much detail as possible, that is, it needs to be complex enough so that it explains a lot. "She died by violence" is presumably not going to be the best explanation we can find of the death, since it doesn't explain the actual details of the death, and there are natural and important questions left unanswered. How detailed an explanation should be depends a great deal on the context of the discussion. Finding out that she died by violence may be quite significant, and a worthy achievement, if we are dealing, say, with the remains of a prehistoric woman 100,000 years old. If it is one's sister, then we need to know a good deal more—we would like to know who did it, how, when, for what reason. It also, of course, seems likely we *can* know a good deal more. Information that cannot be obtained, or that can only be obtained with an unreasonable investment of resources, is reasonably regarded as information we don't need. The drive in science, of course, is to provide the simplest explanations that explain as much as possible, and the dialectic of science often revolves around the difficulty in meeting these conflicting requirements.

**It should not be too *ad hoc*.** That is, it should explain with as little in the way of additional, special assumptions about the world *not supported elsewhere in our belief system* as possible, so the explanation should make as little in the way of new assumptions as can be managed. An *ad hoc* assumption is tailor made to support our explanation, with no good reason to believe it, and generally good reason not to believe it. One can always explain a murder, for instance, by postulating an assassin nobody knows who knew the victim a long time ago and has a grudge against him for some reason we make up, and used some very complicated method we also make up, and perhaps has special powers, and so on. Every murder could be explained by space aliens with fantastically advanced technologies who cover up their tracks perfectly, making it look like someone else did it. This last is an example of the most effective *ad hoc* move available to a philosopher—design a form of your theory that predicts that there won't be any evidence for it, or even that there will be considerable evidence against it. It enables you to treat every failure to explain as a success, for if your theory is true, it will appear to be false. The space aliens, according to the theory, cover up their tracks perfectly.

To attack an argument to the best explanation, one does well to see if there isn't an explanation of the facts not included in the list the argument works from, an explanation which may turn out better than the one being argued for. Also, of course, one may try to show that the explanation favored in fact does not do as well by these criteria as its proponent thinks, or that

some already known, alternative explanation does better than the argument admits. What facts does the person advancing the argument want to explain? A narrow enough selection of facts to explain can make any explanation look good, so see if there aren't important facts that the explanation would predict should not be facts at all. Also, of course, check the facts. It's remarkable how often people use this form of argument, and it turns out that the facts to be explained are not facts at all. (So one might argue that some theory must be true because it is the best way to explain UFO sightings, or the mysterious disappearances in the Bermuda Triangle).

In general, theories about the world, scientific and otherwise, are accepted on the basis of their explanatory power. Virtually every theory will have a few problems, some facts which it fails to explain, and some generally accepted theories have a lot of problems, but as long as we find it plausible that the problem is due, not to the theory's falsehood, but to our not yet having discovered *how* the theory explains the fact, or to faulty background information, or some such thing, the theory may continue to be well regarded. The question is, can the theory (or some future revision of it) be made to work, so that its current failings are our fault (not enough time put in, miscalculations, failures of insight, and so on), or are the problems the theory's fault. In particular, as long as a theory continues to be theoretically fruitful, and to explain new facts and make new kinds of predictions that work out well, it will continue to be well-regarded. Once it ceases to do this (is no longer "progressive"), then we start casting about for a replacement that will do a better job, and if enough really important unexplained facts accumulate, we will be willing to do a fairly extensive reconstruction of our beliefs to get a better, progressive, theory in its place. It is like someone who owes a lot of money. That does not make the guy a deadbeat, as long as he is paying off his debts steadily and can be expected to pay them off in accord with the schedules agreed on. If we begin to think he cannot or will not pay off his debts on time, or possibly at all, then he will no longer be reasonably regarded as a good credit risk.

Sometimes a new scientific theory will replace an old one, and people might say it explains the facts the old one explained in a new way (and explains a lot of other facts the old one does not explain). When this happens the new theory will often describe the facts to be explained in a new way. Instead of explaining why the girl was susceptible to the bad effects of her sister's hostile spiritual force, we explain why she was susceptible to bacterial infection, say. It is probably better to say that the new theory explains the success of the old one in explaining its "facts," "facts" which the new theory may re-describe or not recognize as facts at all. That is to say, a theory replacing an older theory should explain why the older theory succeeded as well as it did in explanation and prediction. (It might do this by suggesting that the older theory's predictions were approximately true, and true for all we could see, given the range of error in our instruments at the time. So Einstein's relativity theory makes it out that Newtonian mechanics is false, but, at velocities well below the speed of light, if Einstein's theory is correct, it produces predictions so close to those of Newton's theory that it is impossible, given the accuracy of our measurements, to show the Newtonian predictions are wrong. That means that all the experiments supporting Newton at low velocities also support Einstein. It is only by observing things near the speed of light that we can test the two theories against each other and see which one works best. A new theory might also decide that the problem was that a lot of "facts" weren't facts at all, so the treatment of biblical revelation by modern science, which no longer feels a need to take the account of the origin of the world in *Genesis* as factual.) The theory should also, of course, explain more than the old theory, or connect to other fields more fruitfully—it should be superior to the old theory in some way or other.

Scientific theories age just as people do. A new theory will typically have a lot of promise and be very suggestive, but not much of it will be worked out yet. As it matures, most of its failures at explanations will be taken care of, it will produce a number of successful predictions, and suggest a number of interesting theoretical options in other areas. Gradually its fruitfulness will drop off, until it becomes boring and routine, and perhaps more or less indispensable for our work, and as background to the explanations of other theories. This is "normal science," and it seems clear now that much of what we regard as science (the atomic theory of chemical combination embodied in the periodic table of the elements, for instance) may constitute permanent acquisitions to our knowledge. But a theory regarded as routine may also begin to accumulate problems which seem to resist solution (observations that it fails to explain, or which even contradict its predictions outright, poor fit with other parts of our science, and so on). People become less and less happy with it, until an alternative is hit upon which explains the successes of the old theory and solves a lot of its unsolved problems, while making interesting new predictions and suggesting new theoretical approaches elsewhere. This leads to a "scientific revolution" in which the old theory is thrown out in favor of the young new theory, and the whole thing starts over again. (One needs to be careful here. It isn't clear that traditional chemistry is thrown out and replaced with a new theory if we come up with new theories about what is presupposed in traditional chemistry, about the nature of atoms, or the chemical bond, for instance. We have ways of holding on to old, useful theories while expanding the depth and extent of our knowledge. But they don't always work out. The disease theory is not an account of the real nature of spirit possession.) It can happen that an old theory will be resurrected in a new version after a long time, as, perhaps, Atomism, an Ancient Greek notion, was resurrected in the sixteenth century. The "refutations" of the old theory no longer look any good, so we give it another try. A more modern example is the resurrection of the particle theory of light, abandoned in the eighteenth century because of interference phenomenon, and then resurrected in twentieth-century physics. What happened was that the assumption needed for it to explain interference phenomena, the assumption that the laws of motion were different for very small particles from what they were for macroscopic things, had come to be regarded as plausible. That assumption had been suggested in the eighteenth century, but it was reasonably regarded as absurd by the "revolutionaries," and *ad hoc*, there being no reason to suppose it was true at that time other than the fact that it would help out the particle theory of light.

In philosophy it is often claimed that an explanation is the *only possible* explanation of the facts. That should always arouse suspicion. If it is really the only possible explanation, that means it is logically impossible that the facts explained be true and the explanation proposed be false, so we have an attempt here to make an argument to the best explanation deductive. In more ordinary talk, the claim that it is the only possible explanation usually comes to this, "it is overwhelmingly the most probable explanation of all those I can think of, given what we know." But the philosopher often means to say the absurd thing he is committed to if taken perfectly literally, not the reasonable thing we can usually take a person to be intending. Generally the "only possible explanation" turns out to have no empirical content, then, so that any observation one can imagine is consistent with it. To explain is to choose from alternative possibilities, which is why it is generally informative when one provides an explanation. "Ah, *that* (not this or the other thing) is the reason!"