The Best Explanation: Criteria for Theory Choice

Paul R. Thagard

The Journal of Philosophy, Vol. 75, No. 2 (Feb., 1978), 76-92.

Stable URL:
http://links.jstor.org/sici?sici=0022-362X%28197802%2975%3A2%3C76%3ATBECFT%3E2.0.CO%3B2-T

The Journal of Philosophy is currently published by Journal of Philosophy, Inc..

accounts will be sensitive to these objections—that successive theories tend to fail to have the logical relations of contradiction and explanation as a special case or an approximation. Although Carnap does not pursue this all the way to the observational level as Kuhn and Feyerabend do, these problems do arise for him on the theoretical level. But science typically has some cumulative development on this level as well as on the observational one. If these problems are to be avoided, it seems that some noncontextualist link between theoretical terms and the world is needed.

JANE ENGLISH

University of North Carolina

THE BEST EXPLANATION: CRITERIA FOR THEORY CHOICE *

GILBERT HARMAN [1] and others have argued that inductive inference is inference to the best explanation. The major weakness of this claim is the lack of specification of how we determine what hypothesis or theory is the *best* explanation. By what criteria is one hypothesis judged to provide a better explanation than another hypothesis? Except for some very brief remarks about choosing a hypothesis that is simpler, is more plausible, explains more, and is less ad hoc, Gilbert Harman addresses the problem only as it concerns statistical inference.[2] In later work, Harman talks rather vaguely of maximizing explanatory coherence while minimizing change.[3] Keith Lehrer has even remarked upon the "hopelessness" of obtaining a useful analysis of the notion of a better explanation.[4] However, I shall show that actual cases of scientific reasoning exhibit a set of criteria for evaluating explanatory theories. Besides filling in a crucial gap in Harman's account of inference to the best explanation, the criteria furnish a comprehensive account of the justification of scientific theories. I shall argue that this account has many advantages over the hypothetico-deductive model of theory confirmation.

* I am grateful to B. C. van Fraassen, T. A. Goudge, and Dan Hausman for comments on earlier versions.

1 "The Inference to the Best Explanation," *Philosophical Review,* LXXIV, 1 (January 1965): 88–95.

2 "Detachment, Probability, and Maximum Likelihood," *Noûs,* IV, 4 (November 1967): 404–411.

3 *Thought* (Princeton: University Press, 1973), p. 159.

4 *Knowledge* (Oxford: Clarendon Press, 1974), p. 165.

I

The phrase 'inference to the best explanation' is relatively new, but the idea is old. Inference to scientific hypotheses on the basis of what they explain was discussed by such nineteenth-century thinkers as William Whewell and C. S. Peirce, and earlier still by David Hartley, Leibniz, and Descartes. To put it briefly, inference to the best explanation consists in accepting a hypothesis on the grounds that it provides a better explanation of the evidence than is provided by alternative hypotheses. We *argue* for a hypothesis or theory by arguing that it is the best explanation of the evidence.

Inference to the best explanation is common in the history of science. An explicit example of an argument to the best explanation is Charles Darwin's long argument for his theory of the evolution of species by means of natural selection. In his book *The Origin of Species* he cites a large array of facts which are explained by the theory of evolution but which are inexplicable on the then-accepted view that species were independently created by God. Darwin gives explanations of facts concerning the geographical distribution of species, the existence of atrophied organs in animals, and many other phenomena. He states in the sixth edition of this book:

> It can hardly be supposed that a false theory would explain, in so satisfactory a manner as does the theory of natural selection, the several large classes of facts above specified. It has recently been objected that this is an unsafe method of arguing; but it is a method used in judging of the common events of life, and has often been used by the greatest natural philosophers.[5]

Many other quotations could be given to show that Darwin's argument in *The Origin of Species* consists in showing that his theory provides the best explanation.

One of the greatest advances in the history of chemistry was the development by Antoine Lavoisier of the oxygen theory of combustion, which replaced the accepted theory based on the hypothetical substance phlogiston. Lavoisier offered explanations of combustion, calcination of metals, and other phenomena where there is absorption of air. He stated:

> I have deduced all the explanations from a simple principle, that pure or vital air is composed of a principle particular to it, which forms its base, and which I have named the *oxygen principle,* combined with the matter of fire and heat. Once this principle was ad-

[5] *The Origin of Species* (New York: Collier, 1962), p. 476.

> mitted, the main difficulties of chemistry appeared to dissipate and vanish, and all the phenomena were explained with an astonishing simplicity.[6]

According to the accepted phlogiston theory, burning objects *give off* the substance phlogiston, whereas, according to Lavoisier, burning objects *combine* with oxygen. The main point of Lavoisier's argument is that his theory can explain the fact that bodies undergoing combustion increase in weight rather than decrease (625). To explain the same fact, proponents of the phlogiston theory had to make such odd assumptions as that the phlogiston that was supposedly given off had "negative weight." Because the oxygen theory explains the evidence without making such assumptions, it can be inferred as the best explanation.

Other examples of arguments to the best explanation, this time in physics, are to be found in the history of the wave theory of light. In his *Treatise of Light* published in 1690, Christiaan Huygens argued for his wave theory of light by showing how it explains the rectilinear propagation of light, reflection, refraction, and some of the phenomena of double refraction.[7] The wave theory was eclipsed by Newton's particle theory, but Thomas Young attempted to revive the wave theory in three articles published between 1802 and 1804. The main improvement in Young's theory over Huygens' was the addition of the law of interference, which enabled the theory to explain numerous phenomena of colored light.[8] Finally, in a series of articles after 1815, Augustin Fresnel attacked the particle theory by arguing that the wave theory explained the facts of reflection and refraction at least as well as did the particle theory, and that there were other facts, involving diffraction and polarization, which only the wave theory could simply explain. He wrote to Arago:

> Thus reflection, refraction, all the cases of diffraction, colored rings in oblique incidences as in perpendicular incidences, the remarkable agreement between the thicknesses of air and of water which produce the same rings; all these phenomena, which require so many particular hypotheses in Newton's system, are reunited and explained by the theory of vibrations and influences of rays on each other.[9]

Hence the wave theory should be inferred as the best explanation.

[6] *Oeuvres* (Paris: Imprimerie Impériale, 1862), vol. II, p. 623, my translation.
[7] Silvanus P. Thompson, trans. (New York: Dover, 1962).
[8] *Miscellaneous Works,* George Peacock, ed. (London: John Murray, 1855), vol. 1, pp. 140–191; see especially pp. 168, 170, 187.
[9] *Oeuvres Complètes* (Paris: Imprimerie Impériale, 1866), vol. 1, p. 36, my translation.

II

The above arguments exemplify three important criteria for determining the best explanation. By "criteria" I do not mean necessary or sufficient conditions. We shall see that the complexity of scientific reasoning precludes the presentation of such conditions of the best explanation. A criterion is rather a standard of judgment which must be weighed against other criteria used in evaluating explanatory hypotheses. The tensions between the three main criteria will be described below. I call the three criteria *consilience, simplicity,* and *analogy.*

The notion of consilience is derived from the writings of William Whewell.[10] Consilience is intended to serve as a measure of *how much* a theory explains, so that we can use it to tell when one theory explains *more* of the evidence than another theory. Roughly, a theory is said to be consilient if it explains at least two classes of facts. Then one theory is *more* consilient than another if it explains more classes of facts than the other does. Intuitively, we show one theory to be more consilient than another by pointing to a class or classes of facts which it explains but which the other theory does not.

To get a more precise definition, let T be a theory consisting of a set of hypotheses $\{H_1 \ldots H_m\}$; let A be a set of auxiliary hypotheses $\{A_1 \ldots A_n\}$; let C be a set of accepted conditions $\{C_1 \ldots C_j\}$; and let F be a set of classes of facts $\{F_1 \ldots F_k\}$. Then T is consilient if and only if T, in union with A and C, explains the elements of the F_i, for $k \geq 2$.

To get the comparative notion, let FT_i be the set of classes of facts explained by theory T_i. Then we can choose between two different definitions of comparative consilience: (1) T_1 is more consilient than T_2 if and only if the cardinality of FT_1 is greater than the cardinality of FT_2; or (2) T_1 is more consilient than T_2 if and only if FT_2 is a proper subset of FT_1. These definitions are not equivalent, because FT_1 might be much larger than FT_2, while at the same time there are a few elements of FT_2 that are not in FT_1. In other words, it is possible that T_1 explains many more classes of facts than T_2, but that there are still some facts that only T_2 explains.[11] In cases where these two definitions do not coincide, deci-

[10] *The Philosophy of the Inductive Sciences* (New York: Johnson Reprint, 1967), vol. 2, p. 65.

[11] This admits the possibility that one theory can replace another as the best explanation, even if there is no cumulativity of facts explained. See Larry Laudan, "Two Dogmas of Methodology," *Philosophy of Science,* XLIII, 4 (December 1976): 585–597; Laudan's notion of problem solving appears similar to that of explaining classes of facts.

sions concerning the best explanation must be made according to what theory explains the most important facts, or on the basis of other criteria discussed below.

The most difficult feature of the notion of consilience is the notion of a *class* of facts. Whewell also sometimes wrote of *kinds* of facts, but this misleadingly suggests that the problem is ontological. Rather, the problem is merely pragmatic, concerning the way in which, in particular historical contexts, the scientific corpus is organized. The inductive logician must take this organization as given, just as do the scientists whose arguments are studied. Since in general the proponents of competing theories share the same historical-scientific context, they agree on the division of facts into classes. We, like Newton and Huygens, have no difficulty in deciding that reflection and refraction constitute more than one application of the wave theory of light. (I use J. D. Sneed's term 'application' to refer to a class of facts explained by a theory.[12]) On the other hand, we would probably say that the distribution of species of finches and the distribution of tortoises on the Galapagos islands are not facts of different classes and, hence, amount to only one application of the theory of evolution. They both concern geographical distribution in the given region. If Darwin had had any reasons to expect finches to be distributed in a very different way from tortoises, then perhaps the two species could have been counted as different applications. It is notable that, in the passage from the *Origin* quoted above, Darwin uses the "classes of facts" terminology.

Because applications are distinguished by means of background knowledge and historical precedents shared by competing theories, the theories in general agree about the individuation of applications. Sometimes, proponents of a theory will simply ignore one class of fact, as in many phlogiston theorists' refusal to consider the increase in weight of burning bodies. Unexplained facts are neglected by theorists who are more concerned with developing a theory than with criticizing it. But when a new theory comes on the scene and succeeds in explaining what the old one did, as well as facts previously unexplained, then, as a matter of logic, the old theory must attend to the newly explained facts. Additional complications may arise. Investigations by advocates of a new theory may show that the evidence explained by the old theory was faulty. For example, until Darwin, it was generally believed that there was a definite limit to the amount of variation a species could undergo,

[12] *The Logical Structure of Mathematical Physics* (Dordrecht: Reidel, 1971), p. 27.

either under domestication or in nature; Darwin's study of artificial selection refuted this. Darwin's argument in the *Origin*, especially in the middle chapters on objections, also shows the possibility of debate concerning what the applications of a theory are. But, in sum, the lack of precise methods for individuating classes of facts does not vitiate consilience as a criterion for evaluating explanatory hypotheses.

Another way of saying that a consilient theory explains facts of different kinds might be to say that it explains *laws* in different domains. I have not used the notion of law in defining consilience, because not all the facts adduced in favor of theories are laws. Some are: Snell's law of refraction, Lavoisier's law that the increase in weight of a body burned is equal to the loss of weight of the air in which it is burned, and so on. But other facts are more particular: double refraction in Iceland crystal, the perihelion of Mercury, the distribution of fossils in South America. Moreover, in Darwin's case it would often be more accurate to say that the facts are tendencies rather than laws, for example, the affinities between organisms shown by the accepted classification scheme. Accordingly, I shall not adopt the attractive picture of theories achieving consilience by explaining laws.

The historical relevance of the notion of consilience is manifest. Huygens pointed to classes of facts concerning the propagation, reflection, refraction, and double refraction of light. Young expanded the wave theory, and improved the argument for it by adding to the list facts concerning color. Fresnel improved the argument still further by explaining various phenomena of diffraction and polarization. With his work, the wave theory of light became *obviously* more consilient than the Newtonian theory.

Similarly, Lavoisier presented a range of phenomena of combustion and calcination which his theory explained. By virtue of its explanation of the increase in weight of burning bodies, his theory was more consilient than the phlogiston theory. Darwin's theory of evolution was enormously more consilient than the creation hypothesis, as he showed by stating fact after fact which his theory explains but which are inexplicable on the creation hypothesis.

Many other important examples of consilience can be given. An outstanding one is Newtonian mechanics, which afforded explanations of the motions of the planets and of their satellites, of the motions of comets, of the tides, and so on. But the general theory of relativity proved to be more consilient by explaining the perihelion of Mercury, the bending of light in a gravitational field, and

the red shifts of spectral lines in an intense gravitational field. Quantum mechanics far exceeds any competitor in that it provides explanations of the spectral frequencies of certain atoms, of the phenomena of magnetism, of the solid state of matter, and of various other perplexing phenomena such as the photoelectric effect and the Compton effect.

A consilient theory unifies and systematizes. To say that a theory is consilient is to say more than that it "fits the facts": it is to say first that the theory explains the facts, and second that the facts it explains are taken from more than one domain. These two features differentiate consilience from a number of other notions which have been called "explanatory power," "systematic power," "systematicization," or "unification." For example, Carl Hempel has given a definition of "systematic power" which is purely syntactic, and hence much more exact than the above definition of consilience.[13] However, it is not applicable to the sort of historical examples I have been considering, since it concerns only the derivation of sentences formed by negation, disjunction, and conjunction from atomic sentences "*Pa*"; it therefore does not represent the way in which Huygens, Lavoisier, and Darwin systematize by explaining a variety of facts, including those expressed by laws. A more recent construction by Michael Friedman is an attempt to formalize how an explanation provides "unification" by reducing the total number of "independently acceptable" statements,[14] but serious flaws have been found in it by Philip Kitcher.[15]

Behind such attempts is the assumption that explanatory power can somehow be assessed by considering the deductive consequences of a hypothesis. But deductions such as "*A*, therefore *A*," as well as more complicated examples discussed by Sylvain Bromberger [16] and others, show that not all deduction is explanation. Moreover, it is essential to the evaluation of the explanatory power of a hypothesis that what is explained be organized and classified. To take an example from C. S. Peirce: we may infer that a man is a Catholic priest on the basis that the supposition explains such disparate facts as that he knows Latin, wears a black suit and white collar, is celibate, etc. We are not concerned with the explanation of a horde

[13] *Aspects of Scientific Explanation* (New York: Free Press, 1965), pp. 280f.

[14] "Explanation and Scientific Understanding," this JOURNAL, LXXI, 1 (Jan. 17, 1974): 5–19.

[15] "Explanation, Conjunction, and Unification," *ibid.*, LXXIII, 8 (April 22, 1976): 207–212.

[16] "Why-Questions," in Robert G. Colodny, ed., *Mind and Cosmos* (Pittsburgh: University Press, 1966), pp. 92ff.

of trivial facts from the same class, such as that his left pant leg is black, his right pant leg is black, and so on. In inferring the best explanation, what matters is not the sheer number of facts explained, but the variety, and variety is not a notion for which we can expect a neat formal characterization.

So far, I have been discussing a static notion of the consilience of theories, which presupposes that a totality of classes of facts—the total evidence—is given. This is generally how it appears when a scientist presents the results of his/her research. Arguments to the best explanation cite a range of facts explained. But there is also a *dynamic* notion of consilience which must be taken into account in considering the acceptability of explanatory hypotheses.

Whewell's notion of consilience is essentially dynamic. He says: "The evidence in favour of our induction is of a much higher and more forcible character when it enables us to explain and determine cases of a different kind from those which were contemplated in the formation of our hypotheses" (*loc. cit.*). *Dynamic consilience* can be defined in terms of consilience: a theory T is dynamically consilient at time n if at n it is more consilient than it was when first proposed, that is, if there are new classes of facts which it has been shown to explain. It is difficult to state precisely a comparative notion of dynamic consilience. Roughly, T_1 is more dynamically consilient than T_2 if and only if T_1 has succeeded in adding more to its set of classes of facts explained than T_2 has.

Successful prediction can often be understood as an indication of dynamic consilience, provided that the prediction concerns matters with which the theory used to make the prediction has not previously dealt, and provided that the prediction is also an explanation. Successful prediction in a familiar domain contributes relatively little to the explanatory value or acceptability of a theory: one more correct prediction of, say, the position of Mars would be of limited importance to Newtonian mechanics, although it would reinforce the belief that the theory explains facts of that class. In contrast, Halley's use of Newtonian theory to predict the return of the comet named after him was a mark of the explanatory power of the theory, which had not previously been applied to comets. Another example of this kind of dynamic consilience is Young's application of the law of interference to the phenomenon of dipolarization discovered by Arago and Biot.

In the *conservative* dynamic consilience just described, no modification to the theory T or set of auxiliary hypotheses A is needed to explain the new phenomenon. But often a theory will impress by

managing, through a change in T or A, to explain a phenomenon inexplicable by the theory in its original form. An example of this is Fresnel's supposition that light waves are transverse rather than longitudinal, which enabled him to explain the facts of polarization. I shall call this property of a theory—that, by means of modifications of the theory or auxiliary hypotheses, it succeeds in explaining new kinds of facts—*radical* dynamic consilience. The wave theory of light, developing from Huygens to Young to Fresnel, is an excellent example of radical dynamic consilience. There is one obvious danger in expanding a theory to explain a new fact: the value of the expansion is illusory if the change involves merely the addition of an ad hoc hypothesis, that is, a hypothesis that serves to explain no more phenomena than those it was introduced to explain. Accordingly, we must require that the modified theory prove to be *conservatively* dynamically consilient.

What I call dynamic consilience is similar to Imre Lakatos' notion of progressive problemshifts.[17] Both notions serve to represent the way in which a theory gains support by improving over time. The hypothetico-deductive method neglects this dynamic feature of theory evaluation.

To this point, I have been treating consilience as a property of theories, but generalizations can also be inferred as best explanations. How does a statement so prosaic as "All ravens are black" explain different classes of facts? The answer here lies in the familiar problem of the variety of instances. If we wanted to test the claim that all ravens are black, we would not merely check all the ravens in Ontario. Many instances might thereby be collected, but we would receive much better support for the claim if we checked a smaller sample of ravens from different continents, from different climates, and so on. Any pollster knows that it is more important to get a representative sample, stratified so as to get a cross section of the population, than it is to get a very large sample. A major flaw of the hypothetico-deductive model of testing, and also of induction by simple enumeration, is that no account is taken of the variety of instances. On the view of scientific reasoning as inference to the best explanation, the variety of instances is simply a kind of consilience. A generalization $(x)(Fx \supset Gx)$ is consilient if there is variety among the objects a such that the generalization, in conjunction with Fa, explains Ga. The notion of variety is as prob-

[17] "Falsification and the Methodology of Scientific Research Programmes," in Lakatos and A. Musgrave, eds., *Criticism and the Growth of Knowledge* (New York: Cambridge, 1970), pp. 116ff.

lematic as the notion of class of fact or application, but again background knowledge is the key to classification. To test Snell's law, for example, we would measure refraction in different substances, at different temperatures, and so on, but not bother measuring it in different cities, because of our belief that light in New York is no different from light in London. Laws of combustion should be tested with a variety of substances, where it is clear that variety here means, say, both wood and phosphorus, rather than two ends of the same stick.

According to Wesley Salmon,[18] variety of instances is important in that it helps us to eliminate alternative hypotheses; according to Clark Glymour,[19] variety is needed in order to compensate for cases where errors in one or more hypotheses, or in evidence, may cancel each other out. Glymour's point is independent of consilience, but we can incorporate Salmon's insight by noting that one way in which variety helps to eliminate alternative hypotheses is by enabling us to show that one hypothesis is more consilient than the others.

One final remark on consilience. It would appear that the maximally consilient hypothesis or theory is one that explains any fact whatsoever. This would be achieved by sufficient flexibility in the set of auxiliary hypotheses to ensure that any phenomenon could fall under the theory. Lavoisier accused the phlogiston theory of having this property, and psychoanalytic theory is also often subject to the charge of explaining *too much*. We might therefore want to put an upper bound on consilience, requiring that, for a theory to be consilient, it must not only explain a range of facts, but also specify facts it could not explain.[20] This requirement is unsatisfactory, however, because one way in which a theory could satisfy the upper-bound condition is to specify facts in a totally different field; for example, psychoanalytic theory does not explain the price of gold. Moreover, it is quite legitimate to contemplate adjustments to a theory or to its set of auxiliary hypotheses which would enable it to explain any anomaly within its field. After all, we want a theory to be dynamically consilient. The limit to these adjustments depends on the increase in consilience of the theory being offset by a decrease in satisfaction of other criteria, such as precision and simplicity. Simplicity, to which I now turn, is the most important constraint on consilience.

[18] *The Foundations of Scientific Inference* (Pittsburgh: University Press, 1966), p. 131 f.

[19] "Relevant Evidence," this JOURNAL, LXXII, 14 (Aug. 14, 1975), p. 419 f.

[20] Cf. Glymour's third condition of confirmation, *ibid.*, p. 414.

III

Simplicity is most clearly an important factor in the arguments of Fresnel and Lavoisier. The kind of simplicity involved in these cases has little to do with current notions of simplicity based on syntactic or semantic considerations. Rather, simplicity is intimately connected with explanation.

The explanation of facts F by a theory T requires a set of given conditions C and also a set of auxiliary hypotheses A. C is unproblematic, since it is assumed that all members of C are accepted independently of T or F. But A requires close scrutiny.

An *auxiliary hypothesis* is a statement, not part of the original theory, which is assumed in order to help explain one element of F or a small fraction of the elements of F. This is not a precise definition, but examples should help to clarify its intent. In the case of Huygens, T would include such statements as that light consists of waves in an ether, and that light waves are propagated according to Huygens' principle that around each particle in the medium there is made a wave of which that particle is the center. In order to explain the laws of refraction and reflection and other phenomena, Huygens assumes that waves are spherical. But in order to explain the irregular refraction in Iceland crystal, Huygens supposes that some waves are spheroidal. This last assumption, restricted in use to one class of fact, is an example of an auxiliary hypothesis. Similarly, Huygens assumed that the speed of light is slower in denser media, in order to explain Snell's law of refraction. (Newton's explanation of Snell's law assumed that the speed of light is *faster* in denser media.) The assumptions of spheroidal waves and the speed of light were not independently acceptable at the time of Huygens, so they do not belong in C; and they were not used to explain any phenomena besides those mentioned, so they must be placed in A rather than T. One might want to reserve the term 'theory' for the union of T and A, but this would not reflect historical practice, and would blur the real distinction between statements that figure again and again in explanations and those whose use is much more limited.

Now we can say that simplicity is a function of the size and nature of the set A needed by a theory T to explain facts F. This is the main notion of simplicity used by Fresnel and Lavoisier. Fresnel accused the Newtonian theory of needing a new hypothesis, such as the doctrine of fits of easy transmission and easy reflection,[21] for

[21] See Isaac Newton, *Opticks* (New York: Dover, 1952), p. 281.

each phenomenon that it explained, whereas the wave theory uses the same principles to explain the phenomena. Similarly, Lavoisier criticizes the phlogiston theory for needing a number of inconsistent assumptions to explain facts easily explained by his theory. These examples show how simplicity puts a constraint on consilience: a *simple* consilient theory not only must explain a range of facts; it must explain those facts without making a host of assumptions with narrow application.

An *ad hoc* hypothesis is one that serves to explain no more phenomena than the narrow range it was introduced to explain. Hence a simple theory is one with few ad hoc hypotheses. But "ad hocness" is not a static notion. We cannot condemn a theory for introducing a hypothesis to explain a particular fact, since all theorists employ such hypotheses. The hypotheses can be reprehended only if ongoing investigation fails either to uncover new facts that they help to explain, or to find more direct evidence for them, as in Fizeau's observation in the nineteenth century concerning the speed of light. Moreover, an auxiliary assumption will not be viewed as ad hoc if it is shared by competing theories.

This brings us to a comparative notion of simplicity. Let AT_i be the set of auxiliary hypotheses needed by T_i to explain a set of facts F. Then we adjudicate between T_1 and T_2 by comparing AT_1 and AT_2; but how is this done? The matter is not neatly quantitative, since any AT could be considered to have only one member, merely by replacing its elements by the conjunction of those elements. Nor can we use the subset relation as we did in comparing sets of classes of facts explained, because it is quite possible that AT_1 and AT_2 will have no members in common. A qualitative comparison, application by application, must be made. For example, on the issue of the speed of light in different media, there was a stalemate between the wave and corpuscular theories, because the assumptions they make are of a similar kind, and until the mid-nineteenth century there was no independent evidence in favor of either. On the other hand, Newton's theory has at least one auxiliary hypothesis, the "doctrine of fits of easy reflexion and easy transmission," corresponding to which there is no auxiliary hypothesis in the wave theory. Young's principle of interference, which explains the colors of thin plates at least as well as the doctrine of fits, can be considered as part of the theory by virtue of its explanation of various phenomena concerning fringes. Thus the comparative simplicity of two theories can be established only by careful examina-

tion of the assumptions introduced in the various explanations they provide. As has often been remarked, simplicity is very complex.

The above account of simplicity is superficially similar to one recently proposed by Elliott Sober.[22] Sober defines simplicity as informativeness, where a hypothesis H is more informative than H' with respect to a question Q if H requires less extra information than H' to answer Q. He applies this to explanation by saying that an explanation is simpler the fewer the initial conditions required in the deduction of the explanandum from the hypothesis. Thus if explanandum E is deducible from theory T_1 in conjunction only with initial condition C_1, whereas the deduction of E from T_2 requires conditions C_1 and C_2, then T_1 provides a simpler explanation (48f). This has some plausibility, but Sober does not employ the notion of auxiliary hypotheses which, I have argued, is crucial to simplicity. Lavoisier and Fresnel show no concern about syntactic complexity of the explanations given by their opponents: the number of initial conditions required is irrelevant. What matters is the special assumptions made in explaining particular classes of facts. Hence simplicity goes beyond the syntactic notion of informativeness discussed by Sober.

Besides comparing sets of auxiliary hypotheses AT_1 and AT_2, we might also consider judging simplicity by comparing T_1 and T_2. But I can not see how in general this could be done. The number of postulates in a theory appears to have little bearing on its acceptability; all that matters is that each postulate be used in the explanation of different kinds of facts. Perhaps T_1 and T_2 could be compared as to number of parameters or predicates, but the relevance of this is doubtful. However, T_1 and T_2 *can* be compared at another level—ontological economy. Lavoisier suggests that the phlogiston theory is less simple than the oxygen theory, since it assumes the existence of another substance, phlogiston. Similarly, the creation hypothesis is ontologically more complex than the theory of evolution. One might suppose that the wave theory was actually less ontologically economical than the corpuscular theory, since it assumed the existence of the ether, although Newton's theory had its own major ontological assumption—the existence of light particles.

T_1 is more ontologically economical than T_2 if T_2 assumes the existence of entities that T_1 does not. This criterion of ontological economy is subsidiary to those of consilience and simplicity because

[22] *Simplicity* (Oxford: Clarendon Press, 1975).

Occam's razor counsels us only not to multiply entities *beyond necessity*. Necessity is a function of the range of facts to be explained without the use of a lot of auxiliary assumptions. Ontological complexity does not detract from the explanatory value or acceptability of a theory, so long as the complexity contributes toward consilience and simplicity. Lavoisier can be construed as arguing, not that his theory is better because it is more ontologically economical, but that his theory is more consilient and simple than the phlogiston theory, so phlogiston need not be assumed to exist. Hence ontological economy is not an important criterion of the best explanation.

But simplicity, illustrated by the arguments of Lavoisier and Fresnel, *is* important. Theories must not achieve consilience at the expense of simplicity, through the use of auxiliary hypotheses. Inference to the best explanation is inference to the theory that best satisfies the criteria of consilience and simplicity, as well as a third: analogy.

IV

Analogy plays an important part in the arguments of Darwin and the proponents of the wave theory of light. Darwin used the analogy between artificial and natural selection for heuristic purposes, but he also claimed the analogy as one of the grounds for belief in his theory.[23] Huygens, Young, and Fresnel each used the analogies between the phenomena of sound and those of light to support the wave theory of light.[24] However, at first sight analogy appears to have little to do with explanation. Darwin's analogy between artificial and natural selection and Huygens' analogy between sound and light are intended to support the respective theories, but it is not clear how this is accomplished. I shall argue that the analogies support the theories by improving the explanations that the theories are used to give.

Arguments from analogy are commonly represented as follows:

(AA) A is P, Q, R, S.
B is P, Q, R.
$\therefore$ B is S.

We conclude that an object or class B has a property S, on the grounds that it shares a number of other properties with A, which has S. Thus Darwin might argue that, since natural selection is like artificial selection in a number of respects, it too leads to the de-

[23] See Chapter 1 of Darwin, *op. cit.*, and Darwin, *The Life and Letters* (New York: Johnson Reprint, 1969), vol. 3, p. 25.

[24] Huygens, *op. cit.*, p. 4; Young, *op. cit.*, vol. 1, p. 188, 211; Fresnel, *op. cit.*, vol. 1, p. 13.

velopment of species. Huygens might argue that, since light is like sound in a number of respects, it also consists of waves. Now, perhaps arguments like this capture part of the use to which Huygens and Darwin put analogy, but severe problems are caused by the presence of *disanalogies*. Huygens (10) takes pains to point out numerous ways in which sound and light do *not* resemble each other. Most crucially, sound is not propagated in straight lines. In Darwin's case there is also a patent disanalogy: the absence in natural selection of an intelligent being that performs the selection. Yet in neither case does the presence of disanalogies daunt the arguer. But if there are properties T and U which A and B do not share, surely it is not legitimate to conclude that because A and B share P, Q, and R, they also share S. Hence (AA) does not adequately represent the use of analogy in scientific arguments.

A better characterization of analogical inference can be given by using the concept of explanation. Suppose A and B are similar in respect to P, Q, and R, and suppose we know that A's having S *explains* why it has P, Q, and R. Then we may conclude that B *has* S is a promising explanation of why B has P, Q, and R. We are not actually able to conclude that B has S; the evidence is not sufficient and the disanalogies are too threatening. But, the analogies between A and B increase the value of the explanation of P, Q, and R in A by S.[25]

The criterion of analogy makes possible the incorporation of what N. R. Hanson called the "logic of discovery" into the logic of inference to the best explanation. Hanson claimed that there is an autonomous logic of discovery, consisting of arguments that an explanatory hypothesis will be of a certain *kind*, similar to successful hypotheses in related fields.[26] But to say that H is of the appropriate kind is equivalent to saying that it has certain analogies with the successful hypotheses. The examples of Darwin and the wave theorists show that analogies figure in arguments concerning the best explanation. Because analogy is a factor in choosing the best explanation, there is no logic of discovery distinct from the logic of justification.[27] Analogy may be used either to direct inquiry toward certain kinds of hypotheses or to support hypotheses already discovered. Support may thus be gained for hypotheses that are, for ex-

[25] Cf. Peter Achinstein's notion of Analogical-Explanatory Inference, *Law and Explanation* (Oxford: Clarendon Press, 1971), p. 133.

[26] "Is There a Logic of Discovery?" in Herbert Feigl and Grover Maxwell, eds., *Current Issues in the Philosophy of Science* (New York: Holt, Rinehart & Winston, 1961), p. 23.

[27] This is argued at much greater length in my "The Autonomy of a Logic of Discovery," forthcoming in the *Festschrift* for T. A. Goudge.

ample, uniformitarian rather than catastrophist, mechanical rather than teleological, or determinist rather than statistical, as well as to support hypotheses invoking particular mechanisms such as selection and wave propagation.

But there is still more to the matter. Not only does analogy between phenomena suggest the existence of analogy between explanatory hypotheses; it also *improves* the explanations in the second case, because the first explanation furnishes a model for the second one. Explanations produce understanding. We get increased understanding of one set of phenomena if the kind of explanation used—the kind of model—is similar to ones already used. This seems to me to be the main use of analogy in Huygens and Darwin. The explanatory value of the wave hypothesis is enhanced by the model taken over from the explanation of certain phenomena of sound. Similarly, the explanatory value of the hypothesis of evolution by means of natural selection is enhanced by the familiarity of the process of artificial selection. Explanations in terms of the kinetic theory of gases benefit from the mechanical model of billiard balls.

I am not claiming that explanation is reduction to the familiar: scientific explanations often employ unfamiliar notions and introduce entities as peculiar as positrons and black holes. However, other things being equal, the explanations afforded by a theory are better explanations if the theory is familiar, that is, introduces mechanisms, entities, or concepts that are used in established explanations. The use of familiar models is not essential to explanation, but it helps.

V

Thus analogy, like simplicity, turns out to be intimately connected with explanation. Unlike hypothetico-deductive and Bayesian models of theory evaluation, the best-explanation view gives an integrated account of the nature and importance of simplicity, ad-hocness, analogy, and variety of instances. Because it accounts for many different aspects of scientific reasoning and applies to examples from different sciences, we can say with a hint of circularity that the theory of inference to the best explanation outlined above is a highly consilient one.

Inference to the best explanation also represents the importance of competition among theories. Inference to a scientific theory is not only a matter of the relation of the theory to the evidence, but must also take into account the relation of competing theories to the evidence. Inference is a matter of choosing among alternative theories, and we choose according to which one provides the best explanation.

The above differs from most accounts of theory choice in that the emphasis is on pragmatic notions rather than syntactic or semantic ones. Explanation is a pragmatic notion,[28] and so is consilience, since the organization of facts into classes is a matter of historical context. The presence of pragmatic elements does not imply that theory choice is subjective: theory choice is historically relative only in the benign sense that the application of objective criteria such as consilience presupposes a given scientific-historical context. It can be shown that this concern with the pragmatic translates into the avoidance of the notorious paradoxes of confirmation.[29]

Application of the criteria of consilience, simplicity, and analogy is a very complicated matter. Proponents of the hypothetico-deductive method often assume that one measure, such as degree of confirmation, suffices for theory evaluation. But, as Gerd Buchdahl has urged, there are often tensions among the various components of the support for a theory.[30] Consilience and simplicity militate against each other, since making a theory more consilient can render the theory less simple, if extra hypotheses are needed to explain the additional facts. The criterion of analogy may be at odds with both consilience and simplicity, if a radically new kind of theory is needed to account simply for all the phenomena. Capturing the multi-dimensional character of scientific-theory evaluation is yet another virtue of the view that scientific inference is inference to the best explanation.

I mention as a final merit of the above account that it makes possible a reunification of scientific and philosophical method, since inference to the best explanation has many applications in philosophy, especially in metaphysics. Arguments concerning the best explanation are relevant to problems concerning scientific realism, other minds, the external world, and the existence of God. Metaphysical theories can be evaluated as to whether they provide the best explanation of philosophical and scientific facts, according to the criteria of consilience, simplicity, and analogy.

PAUL R. THAGARD

University of Michigan, Dearborn

[28] See B. C. van Fraassen, "The Pragmatics of Explanation," *American Philosophical Quarterly*, XIV, 2 (April 1977): 143–150.

[29] Space does not permit discussion of how inference to the best explanation deals with Hempel's raven paradox and Goodman's "grue" paradox. On the transitivity paradox, see B. A. Brody, "Confirmation and Explanation," this JOURNAL, LXV, 10 (May 16, 1968): 282–299.

[30] Gerd Buchdahl, "History of Science and Criteria of Choice," in Roger H. Steuwer, ed., *Minnesota Studies in the Philosophy of Science*, vol. 5 (Minneapolis: Univ. of Minnesota Press, 1970), pp. 204–230.