Explaination—Opening Address

J. J. C. SMART

It is a pleasure for me to give this opening address to the Royal Institute of Philosophy Conference on ‘Explanation’ for two reasons. The first is that it is succeeded by exciting symposia and other papers concerned with various special aspects of the topic of explanation. The second is that the conference is being held in my old alma mater, the University of Glasgow, where I did my first degree. Especially due to C. A. Campbell and George Brown there was in the Logic Department a big emphasis on absolute idealism, especially F. H. Bradley. My inclinations were to oppose this line of thought and to espouse the empiricism and realism of Russell, Broad and the like. Empiricism was represented in the department by D. R. Cousin, a modest man who published relatively little, but who was of quite extraordinary philosophical acumen and lucidity, and by Miss M. J. Levett, whose translation of Plato’s Theaeetus formed an important part of the philosophy syllabus.

Despite my leaning towards empiricism and realism, I remember having been impressed by Bernard Bosanquet’s Implication and Linear Inference. In the main, one must regard this book as a book not on logic but on epistemology. Linear inference is the sort of inference we get in a mathematical proof, but Bosanquet’s notion of implication is much closer to the contemporary notion of inference to the best explanation, which Gilbert Harman, who introduced the term, has argued to be the core notion of epistemology. I shall in general be agreeing with Harman here, though I shall not take a stand on the question of whether all inductive reasoning can be so described. Bosanquet’s notion of implication stresses the notion of fitting into a system. In this way, also, it is not too far from Quine’s and Ullian’s metaphor of the web of belief. The notion of argument to the best explanation indeed goes back to the idea

1 The Theatetus of Plato. Later reissued with a preface by D. R. Cousin (University of Glasgow Press, 1977). (Originally published by Jackson, Wylie and Co.: Glasgow, 1928.)
of coherence as we find it in the writings of writers such as Bradley, Bosanquet and Joachim, though it is important to stress that coherence should be seen as an epistemological notion, not a semantical one. In other words, coherence provides a bad theory of truth, but a good theory of warranted assertability.5

This brings us to the main theme of this conference. I want to characterize explanation of some fact as a matter of fitting belief in this fact into a system of beliefs. The affinities to the coherence theory and to the notion of the web of belief are obvious. In fact I think it just is the coherence theory, so long as it is remembered that I take the coherence theory to be a theory of warranted assertability, not a theory of truth, an epistemological theory, not a semantical or ontological theory. Of course the notion of fitting into a system or of coherence is not at all a precise one. This is, I think, because the notion of explanation itself is not a very precise one. Nevertheless though the notion of coherence may not be a very precise one, it seems to me to be a philosophically important one, and I think that, as I shall shortly try to explain, it has gained in scientific respectability, with all its imprecision, by being put into relation with concepts and investigations of artificial intelligence, in particular by some work by Paul Thagard,6 and also by recent work in the dynamics of belief systems.7

The concept of explanation is a somewhat polymorphous one, though that is not the source of the sort of imprecision that I had in mind in the last paragraph. Mainly I shall be concerned with scientific and historical explanation i.e. the explanation of particular facts (and in the case of the exact sciences) of laws. There are peripheral uses of the terms 'explain' and 'explanation'. For example, in the previous paragraph I said that I would shortly try to explain something to do with the notion of coherence. This was not a promise to try to give a scientific or historical explanation of the origins of the notion of coherence. Similarly, a teacher might explain to a pupil the meaning of a word. This is also a peripheral sense. Nevertheless even these peripheral cases can perhaps be said to fall under the general notion of fitting something into a web of belief. I class mathematical explanation with scientific explanation. This would be resisted by those who, unlike me, regard the notion of causation as essential to scientific explanation. Thus one might ask for the explanation of the fact that e^a = -1. The explanation would come from the proposition that \( \cos \theta + i \sin \theta = e^{i\theta} \), and the derivation of this last formula. All these cases whether central or peripheral, come under the hospitable notion of fitting into a web of belief.

What explains what? We sometimes say that at a certain time a person explains something to another person. This suggests that an explanation (an explaining) is a speech act. This is how Peter Achinstein has suggested that we think of an explanation.8 On the other hand one may say that the theory of natural selection explains the fact of evolution as discovered from the palaeontological record, or that Newtonian celestial mechanics does not explain the advance of the perihelion of Mercury, while Einstein's general theory of relativity does. This is surely a legitimate notion of explanation, but it could be subsumed under the speech act one by saying that someone's belief in the fact of evolution coheres with his or her belief in Darwin's theory, and so on. C. G. Hempel, with his Deductive-Nomological and Inductive-Statistical patterns of explanation was concerned to explicate our rough notion of explanation in purely syntactical terms. The notion of explanation as a speech act obviously allows us to bring in pragmatic considerations. For McTavish to explain the tides to Jones there must be some considerations about the state of knowledge of Jones. Suppose that Jones already knows the explanation of tides. Does McTavish then explain the tides to Jones? Again suppose that Jones is not competent to follow McTavish's proof that the tidal phenomena follow from laws of nature and statements of initial conditions that Jones accepts or is prepared to accept on McTavish's authority. It would sound a bid odd to say that in this case McTavish succeeds in explaining the tides to Jones.

We might talk of the sentences or propositions asserted by McTavish in his attempt to explain the tides to Jones as an explanation in some sense, akin to Hempel's perhaps, in which an explanation was a linguistic or abstract entity. It is an explanation in the sense that in typical circumstances it would furnish an explanation in the speech act sense.

Talking of \( x \) explaining \( y \) to \( z \) may concentrate attention too much on the didactic type of explanation. Crucially important in the development of science is the discovery of explanations, i.e. cases in which \( z \) is identical to \( x \), as when, for example, Newton learned to explain the

---

5 The term 'warranted assertability' apparently goes back to John Dewey, though Dewey did not make the sharp distinction which I hold that we ought to make, between warranted assertability and truth.


and so on, of doing this. However to explain a fact successfully requires having hypotheses that are acceptable because they fit into one's whole system.

These principles do not of course give us an AI model. Such a model has to be developed by hypothesis and experiment. Thagard has a connectionist model called ECHO. Being connectionist it contrasts with Bayesian models. The model can be thought of as consisting of various neuron like entities connected by various excitatory and inhibitory fibres, but in fact it is represented by a computer programme. Sub-systems of the network can be explanatory, but full explanation comes from the whole network. Now if a hypothesis is in the subordinate way explanatory of observation facts from diverse fields then it possesses consilience. This term was introduced by William Whewell.11 It can be modelled in the network with an inhibitory link from hypothesis 1 to hypothesis 2 if hypothesis 1 explains more pieces of evidence than does hypothesis 2. Simplicity is modelled as follows. Suppose that \( n \) hypotheses together explain a fact \( E \) and \( m \) other hypotheses together explain \( E \), and \( m < n \). Then the excitatory links from \( E \) to the first \( n \) hypotheses will be weaker than the excitatory links from \( E \) to the other \( m \). If a hypothesis \( H \) in the first set is inconsistent with a hypothesis \( H' \) in the second set then \( H' \) will be preferred because it will have more activation from \( E \) than its rival does.

Numbers, or weights, are given to the excitatory and inhibitory links in the network. Typically in Thagard's programme ECHO the excitatory links are given the weight 0.05 and the inhibitory ones are given −0.2.

The idea is that the states of the network will evolve in such a way that with luck it will settle down so that only one hypothesis of any pair of contradictory hypotheses is accepted. So the network will have settled down to accepting the best explanation of the evidence (represented by exciting evidence elements).

Thagard tried his programme ECHO12 successfully on simplified versions of various episodes in the history of science, for example Darwin's theory of natural selection (as contrasted with the theory of special creation of species) and Lavoisier's theory of combustion (as contrasted with the phlogiston theory).13 The computer programmer sees to it that the right answers come out, and so the AI processes model the inductive reasoning (argument to the best explanation) of these

9 Nicholas Rescher in his book *The Coherence Theory of Truth* (Oxford: Clarendon Press, 1973), has worked on revising the coherence theory, though I do not agree that it should be seen as a theory of truth.

10 'Explanatory Coherence', op. cit.


12 As reported in his 'Explanatory Coherence', op. cit.

13 For earlier work see Thagard's *Computational Philosophy of Science*, op. cit.
historic figures. Thagard gives values to the hypotheses and evidence (in fact giving competing hypotheses and different pieces of evidence equal values) and to the excitatory and inhibitory links, and to the effect of simplicity, consilience and analogy. These values are not entirely arbitrary, since they were arrived at by experimenting with what values give good performance. The empirical success of the model therefore suggests that these values correspond to something objective (and the same could be said about the values corresponding to the so-called intuitions of human scientists), even though it is not clear what. This empirical and connectionist approach thus has advantages over the usual Bayesian approach, where probabilities of hypotheses cannot be assigned except in special cases. As Thagard remarks, Bayesian methods work well in diagnostic AI because probabilities can be assigned on the basis of medical statistics. The Ramsey method of determining subjective probabilities in terms of betting quotients cannot be applied: as Thagard says, he would not know how to bet on the theory of evolution. I take the point, but I am not sure how to understand it. I would bet any amount on the theory of evolution (its details apart). How would we collect the bet from a creation scientist? Is the idea that someone before learning about and studying the theory of evolution would bet on the conclusion he or she would come to after learning about it and studying it, thus giving it an initial subjective probability?

I have hopes that this quasi-empirical support will perhaps give some scientific respectability to the more intuitive notions of the coherence theory of warranted assertability. Thus encouraged I shall make use of the admittedly imprecise notions of consilience, simplicity, analogy and fitting into a web of belief, or in short of coherence.\textsuperscript{14}

It was noted that Thagard gives observation reports an independent degree of acceptability. This is necessary, I suppose, for his simplified models, but the full coherence theory need not make this concession to

\textsuperscript{14}This is not of course to decry in any way the efforts of those who have tried (or who do or will try) to precisify these notions. Eminent among them is one of the participants in this conference, Elliott Sober. In his book \textit{Simplicity} (Oxford: Clarendon Press, 1975) Sober defines simplicity as informativeness with respect to a question: a hypothesis is simpler than another to the extent that it requires less extra information for us to answer the question. Sober’s approach is syntactic which in effect is to assume that all atomic propositions are equally informative. Thagard (\textit{Computational Philosophy of Science}, p. 85) remarks that scientists typically are not concerned about this sort of syntactic simplicity, but regard the number of statements of initial conditions required to supplement a theory to derive an observational statement as irrelevant to the simplicity of a theory. A semantic notion of simplicity seems as fraught with as many difficulties as the notion of subjective probability. But all the same, we need both notions.

an atomistic empiricism. It is part of our web of belief that observation is usually reliable. We know about the way vision works, how light rays enter the eyes, are refracted to produce images (in the sense of geometrical optics) on the retina, about the visual cortex, and so on. We also know about illusions and how to try to avoid them. Furthermore the theory of evolution by natural selection leads us to another understanding of why we should expect our eyes, ears and other senses to be on the whole reliable. So the belief that observation statements cohere with the rest of our web of belief means that giving extra weight to observation statements is not really to go outside the coherence theory of warranted assertability.

Similar reflections lead us to understand why the coherence theory should lead us to make fresh observations in order to test our theories. Furthermore one explanation will be a better explanation than another if it also explains a set of phenomena from a different field (consilience). So the requirement that scientists and others consider total evidence available to them and also seek more evidence is taken care of in a way in which it is not taken care of by the simple requirement that the hypothesis should be the most probable (relative to the evidence). That is, not mere self-coherence but \textit{comprehensiveness} belongs to the notion of coherence. The web of belief has its own tendency to expand. We want our beliefs to be true, and truth differs from coherence. We may be warranted in asserting something that is false. Indeed, I think that even a Peircean ideal science might be false in important respects. The universe might trick us. Note that I elucidate this ‘might’ in terms of consistency with core science. There would indeed be a pragmatic paradox (but not a semantic one) in asserting that core science itself might be false. It is part of our web of belief that we are able to weed out false theories by testing as severely as possible the theories we already have. Hence it is part of our web of belief that we can increase the chance of our theories being true, or at least possessing verisimilitude (approximation to truth) if we make further observations and test our theories with respect to them.

In Thagard’s models the observation statements, or what serve as such, are taken to be unproblematic. In real life this is not so, and an AI model to take account of this fact will require new features. There is no bedrock, and explanation is holistic. So sometimes we do not explain facts but explain them away. For example, I am extremely resistant to considering or trying to think up hypotheses that might explain putative paranormal phenomena. Since I cannot fit telepathy or clairvoyance into my web of belief I prefer to put alleged reports of such things down to guilibility, wishful thinking, bad handling of statistics, sheer coincidence, or other ways of resisting belief in paranormal phenomena. The simplest way of fitting the observed phenomena,
which are not the supposed paranormal ones, but the putative reports, into my web of belief is to refuse to take them at face value. We may compare the case of conjurors who have done the most magical looking things in front of our eyes, and in ways of which we do not have the slightest conception. Still, it is part of my web of belief that my eyes can be made to deceive me.

Continually, parts of the web of belief are of course being reconstructed. Generally, though not always, observations that cannot be fitted into the web lead to replacement of one hypothesis within the web by another. Sometimes this can come about by considerations of simplicity and analogy. Considerations of simplicity and analogy may lead us to replace one hypothesis within the web by another. In Thagard's simulations the set of alternative hypotheses is given, but in real life new hypotheses get thought of by the not very well understood process that C. S. Peirce called 'abduction'. The special theory of relativity has received good confirmation from the Michelson–Morley experiment and with observations of fast particles, so that one could not design a cyclotron without using the dynamics of special relativity. Nevertheless Einstein was motivated to a large extent by considerations of simplicity and analogy. Until relatively recently the general theory of relativity had very little, if any, hard observational or experimental support. Einstein and others accepted it largely on the sorts of a priori considerations that I have been mentioning.

Thus when I talk of fitting something into a web of belief, the web must not be thought of as something static into which the new observation reports or sub-theories must fit. Sometimes the fits come by ripping apart large chunks of what had seemed to be a strong and secure region of the web. How beliefs should be accepted, rejected, strengthened or weakened is a matter for the dynamics of belief. It is an open question how precise and how objective such a study may be. We do need such a study to improve on the metaphor of 'fitting in' to a web of belief or the turn of the century talk of 'coherence'. I have referred to the work of Thagard on AI modellings of scientific controversies as pointing to one way of getting objectivity into a study of argument to the best explanation. Another line of research which may go in the same general direction is the work on the statics and dynamics of belief systems by Brian Ellis, Peter Gärdenfors and Peter Forrest. As with the case of the AI modelling, I am not very knowledgeable about this line of research and I shall have to remain on a more imprecise and elementary level.

It is not surprising, perhaps, that such a vague theory of explanation as that of coherence can incorporate the insights of various well known accounts of explanation.

Explanations, like screwdrivers, can be good and bad. Indeed the notion of 'argument to the best explanation' implies that some explanations can be better than others. Note here that 'best explanation' here really means 'best hypothesis'. (It can hardly mean 'best explanatory speech act'.) However the term 'argument to the best explanation' indicates in what way the hypothesis is best, namely that it is the hypothesis belief in which fits best into our web of belief. This contextuality is taken care of by the conception of explanation as a speech act. Not only can explanations be better than others, but we can surely talk of explanations as being positively (if that is the right word here) bad. An explanation may be bad if it fits only into a bad web of belief. It can also be bad if it fits into a (possibly good) web of belief in a bad sort of way.

Thus a creation scientist's web of belief is a bad one. Essential to his or her web is a belief that Genesis and other books of the Bible are the literal word of an infallible Deity. Grant this proposition and no doubt the rest follows easily. However this proposition is a highly un plausible one and the best hypothesis surely is that Genesis, etc. is based on even earlier myths and on the ideas of ancient people with none of the advantages of modern cosmology who were doing the best they could. Modern cosmology, together with evolutionary biology, provides a more comprehensive web of belief. Moreover the creation scientist's total theory is far from simple: it requires a separate creation hypothesis for each species of plant and animal. It is no good just referring to the will of God, because it is still unexplained why God made the various decisions in the different particular cases.

An explanatory hypothesis may be bad if it fits into a good web of belief in a bad sort of way, e.g. to a wrong analogy, a mathematical mistake, a wrong assessment of plausibility, and so on.

One way in which an explanatory hypothesis may be best is of course that it be true. As critics of the correspondence theory of truth have pointed out, we cannot directly compare a hypothesis with facts in the world. This is a poor criticism of the correspondence theory of truth, for why should an adherent of such a theory (or similar theory) wish to say that we could make such a direct comparison? Though a poor criticism of the correspondence theory of truth, it does make a point.

15 Peter Forrest prefers to speak of 'argument to the best theory' instead of 'argument to the best explanation'. I partly agree, but still prefer 'argument to the best explanation', because it is the best theory that provides the best explanation, and it is best because it provides the best explanation for the person who wants the explanation. People with different presuppositions, whose puzzles or 'why?' questions are different, might require different hypotheses to provide an explanation for them. The hypothesis is not a speech act, but the explanation is, even when it is talking to oneself.
about warranted assertability. We cannot directly compare hypothesis with fact. The best we can do is to choose the best warranted hypothesis (in this sense the best explanation) in the hope that warranted assertability is a pointer towards truth. The pointer may lead us astray, since the universe may always trick us. Still, we do the best we can, which is pretty well. At least it is pretty central to my web of belief that I do pretty well. (For example evolutionary theory is part of my web of belief, and this makes it most unpalausible that my senses are very unreliable.)

So when we choose the best explanation, the criteria for 'best' do not include 'true'. In the AI model the best hypothesis is the one that settles down as best on the operation of the network. In a wider conception, allowing for possible rejection of observation reports, it is the one which survives in the modified web. Moreover it is the best of the alternative explanatory hypotheses that occur to us, or that the network is able to add to its store by an AI process analogous to Peircean abduction. It is too much to ask that it be the best of all possible alternative hypotheses, laid up in some platonic or set theoretic heaven. We do the best we can, and we cannot consider hypotheses of which we cannot think.

The history of science suggests that theoretical assertions often get replaced by others with greater verisimilitude. Not always. I think that there are many hypotheses in science that will never be improved on. Will it ever be denied that the lightest isotope of hydrogen contains one proton and one electron? However much we change our conception of electrons and protons still this proposition about hydrogen will hold up. Generalizing, there is a vast reservoir of accumulated scientific fact and theory that will never be overturned. We can recognize Kuhnian revolution without jettisoning much of the cumulative or 'Whig' conception of scientific advance. Still, in fundamental matters we hope for verisimilitude rather than truth. Now a better model of the web of belief would have to take account of the fact that in the face of new evidence or of new serendipitous abductions a good web settles down to a state that does not so much probably maximize verisimilitude, as (in terminology analogous to that of expected utility of economists and utilitarian moralists) maximizes the expected verisimilitude. (In saying this, I am assuming that good sense can be made of the concept of verisimilitude. That is a big 'if' and yet the concept of verisimilitude seems to me to be indispensable to much philosophy of science and metaphysics.16)

16 For work on verisimilitude and for references to other recent work on the subject see Graham Oddie, Likeness to Truth (Dordrecht: D. Reidel, 1986). The subject is a highly technical and contentious one, and yet if it can be brought to fruition the work on it is surely of the greatest importance for the philosophy of science.

Vague and imprecise as I must admit that the characterization of explanation as fitting the explanandum proposition into a system, is this very vagueness and imprecision allows me to subsume many of the well known theories of explanation as special cases.

We may begin with the notion that explanation is a matter of reduction to the familiar. As a general characterization of explanation this is clearly quite wrong. When the facts of black body radiation did not fit in into the webs of belief of classical physicists, Planck introduced the notion of quanta, which affected parts of the web which then had to be repaired in quite novel ways. The history of science suggests that most often explanation is reduction to the unfamiliar. Of course good popularization of novel scientific theory, while stressing the novelties and unfamiliarities, does try, so far as possible, to make the theories acceptable to the intended readers by making copious use of common sense analogies. Even in technical science reduction to the familiar has at least played some part in explanation. Consider, for example, the early kinetic theory of gases, in which we find the gas laws explained in terms of the collisions of small elastic particles. Something like the notion of explanation as reduction to the familiar was prominent in N. R. Campbell's philosophy of physics.17 Campbell held that good theories could be distinguished from absurdly trivial ones by the presence of analogy, and the analogy of gas particles with such things as billiard balls, provided an example. In other cases, such as that of the Fourier theory of heat conduction, Campbell saw a mathematical analogy between the form of the laws of heat conduction and the empirical laws that the theory is intended to explain. Campbell remarked that Fourier's equation is developed by considering the flow of heat through an infinite plane parallel slab. In quantum mechanics we could point to frequent analogies between Hamilton equations with operators and Hamilton equations in classical mechanics. (In opposition to this, it could be held that the use of such analogies is heuristic only and that the analogies do not add to the strength of an explanation.)

The presence of analogy is not quite the same as that of reduction to the familiar. Campbell could allow analogy even in the case of very unfamiliar looking and esoteric theories, there being analogical resemblance between successive theories but little or none between the first and last theories in the chain. On the other hand, in modern physics we may find little or no analogy between successive theories and I suppose that there is no analogy in the case of the replacement of Aristotle's mechanics by Galileo's. Looking at the matter from the

point of view of Thagard's AI investigations we can say that Campbell's theory works well when the excitatory analogical links are set high.

Indeed in fact Campbell does not see all explanation as reduction of the unfamiliar to the familiar.\textsuperscript{18} He distinguishes two sorts of explanation, one being substitution of the familiar for the unfamiliar and the other being substitution of the simple for the complex. A given explanation can be of both sorts, but Campbell agrees that the latter plays the more important part and is characteristic of scientific explanations. As has been pointed out, simplicity plays a part in Thagard's models.

Indeed it is interesting that there are things that are difficult to explain and yet which are very familiar to common sense and infant physics. We all know that glass transmits light whereas wood does not, but not everyone knows why. Ohm's law is familiar to those beginning to study physics, but its explanation requires the unfamiliar and sophisticated ideas of solid state physics.

In mythology, in the traditional legends of the Australian aborigines, we come across what seems in part to be meant as explanations, though there may be other purposes, in that the stories of a mythical past may have expressed wishes and desires that were unable to be realized in the people's own contemporary desert environment.\textsuperscript{19} Consider the mythical explanations of land forms, which are put down to the activities of totemic ancestors. Thus in Aranda tradition a gap in a range of hills is the remains of a broken fish weir, and cracks and water gutters were supposed to have been scoured out by the body of a great and hideous water serpent.\textsuperscript{20} This represents a sort of fitting into a web of belief: beliefs about present land forms are fitted into beliefs about the activities of water serpents and of people who construct fish weirs. To the scientific eye the explanations are of course extremely unsatisfactory. The web of belief is extended at one point at the cost of irreparable damage and loss at other points. The activities of humans and other animals were so familiar that it was assumed that they did not need explanation. This assumption was mistaken. Indeed the human or animal brain is a horrendously complex entity, whose workings are still very imperfectly known, even though modern neuroscience is taking its first tentative steps along the way. On rare occasions also the myths may speak of mountains or formless plains arising by their own volition.

Of course the supposed mythical events are larger than life and so are unfamiliar, but the concepts used, 'water snake', 'fish weir' and so on, are familiar enough. Even the notion of 'turning into' is a common sense one. Caterpillars turn into butterflies. Children find no difficulty with stories in which a prince turns into a frog. Perhaps such ideas are reinforced by dreams. So the idea of a totemic ancestor turning into the totemic animal, which then solidifies into a mountain range, does not use unfamiliar concepts, even though the supposed events are of unfamiliar sorts. In modern times we have come to see that the messy patterns of common sense events can be simplified only by the replacing of familiar concepts by unfamiliar ones. Thus nowadays an unfamiliar concept such as that of 'parity' does not present itself as puzzling, whereas the familiar concept of volition may raise various sorts of difficulties.

It is no wonder, therefore, that reduction to the familiar does not figure much in contemporary philosophizing about explanation. Some good explanations are of this sort and more are not. Bad and unscientific explanations are quite commonly of this sort.

A very influential theory of explanation has been C. G. Hempel's Deductive-Nomological and Inductive-Statistical patterns of explanation. The deductive-nomological pattern is that of an argument in which we deduce a statement concerning a particular event or else a statement of law. I shall concentrate on the former. The Laplacean ideal was that of an infinite intelligence which was possessed of knowledge of the whole state of the universe at a given instant and which could predict or retrodict any future or past event. In practice of course we have to concern ourselves with finite systems that we can treat as closed from outside influences, which in Newtonian mechanics would queer the pitch if they entered the system from outside with action at a distance (perhaps from infinity), between the time of prediction or retrodiction and the event to be explained. We must also neglect Gödelian possibilities that the statement of the event might be true but not deducible from the premises.\textsuperscript{21}

The Hempelian deductive-nomological model of explanation clearly fits in well with the notion of explanation in terms of coherence. One way of fitting a belief into a system is to show that it is deducible from other beliefs. However, as is well known, Hempel's model suffers from counter-examples, examples for which we need to consider also whether the coherence account does or does not avoid them. Hempel has of course reacted to these counter-examples in his own way, but I shall here simply assume that they are damaging to the deductive-nomological account of explanation as a general definition of explanation (apart from Hempel's Inductive-Statistical pattern which I shall discuss separately). Thus consider Wesley Salmon's example 'No men

\textsuperscript{18} Campbell, Foundations of Physics, 114.


\textsuperscript{20} Ibid.

\textsuperscript{21} This is one reason why the Laplacean syntactical definition of determinism is better replaced by a model theoretic one.
who take the birth control pill become pregnant, Jones is a man who regularly takes his wife's birth control pills, therefore Jones does not become pregnant.22 The major and minor premises are true and the conclusion follows from it. Yet the putative explanation of Jones' non-pregnancy is at least misleading, because it suggests that taking the birth control pill had something to do with Jones not becoming pregnant. It is worse than misleading if it leads Jones to think that if he stopped taking the birth control pills he would very likely become pregnant. The putative explanation does not conduce to coherence of Jones's set of beliefs because it invites adding to his web of belief the proposition that at least some men who do not take the pill become pregnant, which is contrary to common experience, as well as to anatomy and physiology. (I shall ignore the case of the man with the XXY chromosome who had a baby and later became officially a woman. To avoid such counter-examples we would have to define 'man' more narrowly, perhaps as persons with XY chromosomes.)

A clearer problem arises over Sylvain Bromberger's early and well known example of the flag pole.23 Consider a flagpole 20 metres high distant from an observer with his eye at ground level, distant 40 metres from the base of the flagpole. By the laws of Euclidean geometry and simple trigonometry we deduce that the angle subtended at the observer's eye is 26° 31'. Now it is intuitive that the height of the flagpole explains the fact of its subtending an angle of 26° 31' at ground level 40 metres away but that the converse does not hold. Nevertheless, each fits Hempel's pattern. I think that the coherence theory suggests an answer to the asymmetry. The proposition that the flagpole is 20 metres high fits naturally into a system of beliefs about the purposes of flagpoles, how they are made from trees and made a certain length in order to subserve such purposes. Also beliefs about how you would make a flagpole to be 20 metres high. Of course this would ipso facto make a flagpole that subtends 26° 31' 40 metres away at ground level, but the measurements that would need to be made would, except in bizarre circumstances, be of the length of the flagpole. Of course you could imagine a person measuring off 40 metres and then with a theodolite finding a suitable spot on the flagpole to cut it, in which case it would be natural to say that the flagpole was 20 metres high because it subtended the appropriate angle 40 metres away at ground level. Since such circumstances would be very unusual we naturally say that the flagpole subtends the angle because of its height and not vice versa. See also Bas van Fraassen's amusing story24 explaining why a tower was a certain height by the fact of the shadow being a certain length. It is an ingenious story, and in less out of the way cases we of course expect to explain the length of the shadow by the height of the tower, not the other way around.25

Why is the light from distant galaxies shifted to the red? Answer: the recession of the galaxies as the universe expands. We are puzzled by the red shift and the hypothesis of the expansion of the universe is what best enables us to fit the fact of the red shift into a web of belief about the wave nature of light, the Doppler shift and so on. The theory of the expansion of the universe renders the red shift no longer puzzling, whereas the expansion of the universe is hardly rendered less puzzling by facts about red shifts. The pragmatic aspect of explanation renders the asymmetry of 'fitting in' or coherence understandable.

Some philosophers may well say here that I could save myself a lot of trouble by defining explanation in terms of causality. For reasons of my own, I am inclined (unfashionably these days though it was fashionable enough fifty or more years ago) to argue that the notion of causality is unimportant in science, or at least in theoretical physics. Certainly I do not want the notion of causality if it has to be defined in terms of counterfactual conditionals. Counterfactual conditionals are sometimes elucidated, notably by David Lewis, as about possible words other than the actual one.26 I hold that science and metaphysics should concern themselves only with the actual. I would give a metalinguistic account of counterfactuals, and I hold that science and metaphysics should not concern themselves with the metalinguistic (save as a linguistic phenomenon).27

Frank Jackson regards causality as an unanalysable theoretical notion, which arises postulationally by the usual hypothetico-deductive method. This certainly seems to me to be an improvement on defining causality in terms of counterfactuals. Salmon's notion of causality is an

25 Philip Kitcher and Wesley Salmon have critically discussed van Fraassen's application of the story in their paper 'Van Fraassen on Explanation', *Journal of Philosophy* (1987), 315–30, especially 317. They think that van Fraassen brings in extraneous considerations. I am inclined to think that the coherence account, for which no considerations need be extraneous, may avoid these criticisms.
27 Some of the ways of constructing what Lewis calls 'ersatz possible worlds' are essentially metalinguistic. I have not space here to discuss the relevance of ersatz worlds that are supposed to exist in the world of set theory.
star we can regard the relevant part of Peter's world line as consisting of the two equal sides $AB$ and $BC$ of an isosceles triangle, while the relevant part of Paul's world line is $AC$. And since $AB$, $BC$ and $AC$ are all time-like lines, the geometry of Minkowski space-time ensures that $AB + BC$ is less than $AC$. (Analogous to Euclidean geometry, in which two sides of a triangle are greater than the third side. The difference is in the minus signs in the metric.) Now this geometrical explanation gives far greater insight than does an operationalist treatment in the manner of Einstein's (pre-Minkowski) 1905 paper. (Just as Minkowski's exhibition of the Lorentz transformations of special relativity as simply a rotation of axes in space-time gives more insight than does an operationalist treatment of special relativity.) If there are worries about Paul's periods of acceleration and deceleration it can be remarked that the calculation could have taken these into account with the aid of a little differential calculus. If it be said that the real explanation is causal, because the accelerations and decelerations were caused by rocket motors, I reply that does not give the main insight. Let $AB$ be part of the world line of a clock that passes a clock part of whose world line is $AC$ and let $BC$ be a clock which passes the $AB$ clock at $B$ and the $AC$ clock at $C$. Then we explain why the sum of the time differences registered by the $AB$ clock and by the $BC$ clock is less than the time difference registered by the $AC$ clock. No acceleration, and no causes here.

Salmon's philosophical approach seems to fit rather well with quantum field theory in which forces (other than gravitational ones—though there have been speculations about gravitons) are seen as exchange forces. Two particles are attracted to one another as another particle is exchanged between them. (This is a little counter-intuitive. If you are standing on smooth ice and I throw a cricket ball for you to catch, we slide apart, not towards one another!) So leaving Bell inequality or Einstein–Podolsky–Rosen considerations out of it, most of fundamental physics (outside the general theory of relativity) does fit Salmon's picture of the world as a network of causal processes (in his own harmless sense of 'causal').

Salmon sees explanation not as argument but as the giving of relevant information. This information can be statistical in nature as indeed it must be in quantum mechanics. Salmon is thus able to deal with statistical explanation of particular events in a better way than Hempel does. In Hempel's deductive-nomological pattern of explanation we have argument, but there is some uncertainty in his presentation of the 'inductive-statistical pattern'. Thus in the deductive-nomological pat-

---

28 Salmon, ibid., 170ff.
29 And extended to general relativity by subsequent writers. See Adolf Grünbaum, Philosophical Problems of Space and Time, 2nd edn (Dordrecht: D Reidel, 1973), 735–50. For A. A. Robb, see his Geometry of Time and Space (Cambridge University Press, 1936).
tern, Hempel writes the conclusion of the argument under a single line, whereas in the inductive-statistical case he uses a double line and a 'probably' in brackets. This makes the so called argument look like a metalinguistic statement of a relation of conclusion to premises, which is of course consonant with the fact that the validity of inductive arguments is affected by addition of premises, which is not the case with deductive arguments.

Hempel says that a statistical explanation of a particular event is a good one only if the occurrence of the event is given a high probability. Salmon, denying that an explanation is an argument, is able to say that if there is a 1/10 possibility of a radio-active atom emitting an alpha particle in a certain period of time, and the alpha particle is emitted, the probability statement gives a perfectly good explanation of the event. Quantum mechanics assures us that there are no ways of describing the atom which would give us a better probability. Salmon's denial that explanations are arguments works well in this sort of context. What would we say if the probability predicted by quantum mechanics was not 1/10 but $1/10^{1,000,000}$? In this case if I observed the event I would not try to fit the observation statement into my web of belief: I would suppose that my eyes had deceived me or that there was something funny about the experiment. To speak a bit in an Irish manner, I would fit the observation into my web of belief by not believing it. But if the probability were 1/10 I would be perfectly happy.

Salmon describes his account of explanation as objectivist and ontological rather than epistemological. Explanation describes how events fit into an objective causal network. My own account is epistemological, we explain $x$ to $y$ by means of fitting $x$ into $y$'s web of belief. Nevertheless, in our explanation we aspire to truth—we want a true web of belief, and in this respect there is not too much difference between the epistemological and the ontological concepts of explanation. Where I can talk of bad explanations as consisting of the bad fitting of beliefs into a pre-existing web, or of good fittings into a bad pre-existing web, the ontological conception can allow us to talk in terms of descriptions and misdescriptions of the causal network.

The main theories of explanation, then, seem at least to fit into the general coherence theory of explanation as special cases. The lack of general agreement among philosophers of science despite the brilliant, and often painstaking work of leading workers in the field suggests to me that the concept of explanation may be what Wittgenstein called a 'family resemblance' one. This would make it at least highly disjunctive, and I suspect also that even if we confine ourselves to good scientific explanation, it may also exhibit what Waismann called 'open texture', in that patterns of discourse may be proposed in science which we would immediately recognize as explanatory, even though we could not have predicted them in advance. The coherence account of explanation would surely fit anything of the sort that will arise, but I am only too aware that it may protect itself in the future as it may have done in the past only by its own woolliness, as a sheep does on a frosty night. (However, I have suggested that Paul Thagard's AI approach may show promise of acting a bit like a sheep shearer and reducing the woolliness.)

Finally, why do philosophers need a theory of explanation? As we all know there are some who will analyse any concept at the drop of a hat. Then there are those who see explanation as a key concept in epistemology, and so need a theory of it. For me, epistemology is the handmaiden of metaphysics, and since argument to the best explanation is important in the defence of realism, I would like to have a good account of what explanation is. In the face of the brilliant, lucid, and often painstaking work of Hempel, Ernest Nagel, Salmon, Achinstein, van Fraassen and others, all I have had to offer are some imprecise thoughts.

Fortunately most of the symposia to follow in this conference are concerned with explanation in the special sciences, Physics, Biology, Psychology, and the Social Sciences. Here I think that the issues concern the structure of these sciences as much as the nature of explanation itself. From these discussions I hope to learn much, as well as from the appropriately final symposium on the Limits of Explanation.  

---

31 I wish to thank Paul Thagard for kindly commenting in correspondence on a draft of this paper, in particular on suggesting corrections to those pages in which I expound his views. He is of course not responsible for defects that remain, and I hope that readers may be encouraged to read his 'Explanatory Coherence' and his *Computational Philosophy of Science*, already cited.
Explanation and its Limits

ROYAL INSTITUTE OF PHILOSOPHY LECTURE SERIES: 27
SUPPLEMENT TO PHILOSOPHY 1990

EDITED BY

Dudley Knowles

1990

CAMBRIDGE UNIVERSITY PRESS

CAMBRIDGE  NEW YORK  PORT CHESTER  MELBOURNE  SYDNEY