WHAT GOOD IS AN EXPLANATION?

Peter Lipton
University of Cambridge

Introduction

We are addicted to explanation, constantly asking and answering why-questions. But what does an explanation give us? I will consider some of the possible goods, intrinsic and instrumental, that explanations provide.

The name for the intrinsic good of explanation is `understanding', but what is this? In the first part of this paper I will canvass various conceptions of understanding, according to which explanations provide reasons for belief, make familiar, unify, show to be necessary, or give causes. Three general features of explanation will serve as tests of these varied conceptions. These features are:

a) the distinction between knowing that a phenomena occurs and understanding why it does;

b) the possibility of giving explanations that are not themselves explained;

c) the possibility of explaining a phenomenon in cases where the phenomenon itself provides an essential part of the reason for believing that the explanation is correct.

There are many other aspects of our explanatory practices that a good account of explanation and understanding should capture, but these simple tests provide surprisingly effective diagnostic tools for the evaluation of broad conceptions of the nature of understanding. It will turn out that the causal conception of understanding does particularly well on the tests, though of course it too faces various difficulties. The balance of this essay focuses on the causal conception. After addressing some of the difficulties it faces, I will ask

why causes explain. Why, in particular, do causes rather than effects explain? One possible answer is that causes `make the difference' between the occurrence and non-occurrence of what they explain. Several features of our explanatory practices will be adduced to evaluate this hypothesis. In the final section, I will consider an instrumental good of explanation revealed by the account of Inference to the Best Explanation: explanation is an important route to the discovery of causes. This allows a functional explanation of explanation, according to which the question, `Why do causes explain?' may itself have a causal answer.

Three Features of Explanation

There are some simple and relatively uncontroversial features of explanation that can be used to test conceptions of understanding. I will use the three I have just listed. The first of these is the **gap between knowledge and understanding**. Knowing that something is the case is necessary but not sufficient for understanding why it is the case. We all know that the sky is sometimes blue, but few of us understand why. Typically, when people ask questions of the form `Why P?', they already know that P, so understanding why must require something more than knowing that. If one's aim is to get a grip on the goods that explanations provide, it is useful to ask: what more than knowledge does understanding require? And if an account of understanding is unable to make the distinction between knowing that and understanding why, it is a bad account.

The second test feature is the **why regress**. As most of us discovered in our youth and to our parents' consternation, whatever answer someone gives to a why-question, it is almost always possible sensibly to ask why the explanation itself is so. Thus there is a potential regress of explanations. If you ask me why the same side of the moon always faces the earth, I may reply that this is because the period of the moon's orbit around the earth is

the same as the period of the moon's spin about its own axis. This may be a good explanation, but it does not preclude you from going on to ask the different but excellent question of why these periods should be the same. For our purposes, the salient feature of the why regress is that it is benign: the answer to one why question may be explanatory and provide understanding even if we have no answer to why-questions further up the ladder. This shows that understanding is not like some substance that gets transmitted from explanation to what is explained, since the explanation can bring us to understand why what is explained is so even thought we do not understand why the explanation itself is so. Any account of understanding that would require that we can only use explanations that have themselves been explained fails the test of the why regress.

The final feature that I will use to test conceptions of understanding is the phenomenon of what are known as **self-evidencing explanations** (cf. Hempel 1965, 370-374). These are explanations where what is explained provides an essential part of our reason for believing that the explanation itself is correct. Self-evidencing explanations are common, in part because we often infer that a hypothesis is correct precisely because it would, if correct, provide a good explanation of the evidence. Seeing the disemboweled teddy bear on the floor, with its stuffing strewn throughout the living room, I infer that Rex has misbehaved again. Rex's actions provide an excellent if discouraging explanation of the scene before me, and this is so even though that scene is my only direct evidence that the misbehaviour took place. To take a more scientific and less destructive example, the velocity of recession of a galaxy explains the redshift of its characteristic spectrum, even if the observation of that shift is an essential part of the scientist's evidence that the galaxy is indeed receding at that the specified velocity. Self-evidencing explanations exhibit a kind of circularity: H explains E while E justifies H. As with the why regress, however, what is salient is that there is nothing vicious

here: self-evidencing explanations may be illuminating and well supported. Any account of understanding that rules them out is incorrect.

Five Conceptions of Understanding:

Reason, Familiarity, Unification, Necessity and Causation

We now have three important features of explanation: there is a distinction between knowing that and understanding why, the why regress is benign, and good explanations may be self-evidencing. Armed with these test features, I want now to consider five broad conceptions of understanding, conceptions of what intrinsic goods an explanation provides. The first two of these conceptions -- reason and familiarity -- make understanding fundamentally an epistemic matter; the last two -- necessity and causation -- make it metaphysical or ontological. The middle conception -- unification -- can go either way, depending on how it is itself analysed. These conceptions of understanding are not mutually exclusive, because different explanations could provide different types of understanding and because a single explanation could yield more than one good.

First, we have the **reason** conception of understanding. Understanding is here identified with having a good reason to believe. We understand why something occurred when we have a good reason to believe that it did in fact occur, and this good reason is just what an explanation provides (cf. Hempel 1965, 337, 364-376). This view has some attractions. When we ask why-questions, sometimes what we really want is not an explanation but a reason for belief. Here `Why P?' is short for `Why should I believe that P?'. The reason conception provides a unitary account of explanation seeking and reason seeking why questions: both are actually reason seeking. Indeed the word `reason' itself has just this ambiguity: it may mean either reason for belief or reason why. Another attraction of the

reason conception of understanding is negative: it avoids any dubious metaphysical notions, and relies on a notion -- reason for belief -- that we must appeal to in any event if we are to do epistemology at all.

Now for the bad news. The reason conception of understanding may fail all three of our tests. First, it does not adequately distinguish between knowing that and understanding why. In many cases (at least), to know that something is the case requires having reasons to believe it, so if the reason conception were correct, all these would also be cases of understanding why. But this is not so: there are many things we have reason to believe occur and know to occur, yet we do not understand why they occur. Given her expertise and honesty, the fact that your computer advisor tells you that your hard disc is severely fragmented gives you an excellent reason to believe that your hard disc is indeed severely fragmented; but it gives you not the slightest inkling *why* your disk is fragmented. Having a good reason to believe P is clearly not sufficient for understanding why P.

The reason conception is also under threat from the why regress. In one sense of reason, H does not provide a reason to believe E unless there is also a reason to believe H. On this construal of the notion of a reason, the reason conception of understanding would then entail that H can only explain E if H has itself been explained. What the why regress shows, however, is that H may explain E even if H is not itself explained. Finally, the reason conception does not readily account for self-evidencing explanations. If E is a reason for H, H cannot be a reason for E. If the spectral red shift is our reason for believing that the galaxy is receding, then the recession does not provide a reason for believing that the spectrum is shifted: this would be a vicious circle. So if the reason conception were correct, no self-evidencing explanations would be legitimate; but many are.

I turn now to my second contestant, the **familiarity** conception of understanding. This is the view that explanation is in some sense `reduction to the familiar'. It is what is strange or surprising that we do not understand; a good explanation gives us understanding by making the phenomenon familiar, presumably by relating it to other things that are already familiar (cf. Hempel 1965, 430-433; Friedman 1974, 9-11). Loose though this specification is, it is enough to suggest that the familiarity conception of understanding, unlike the reason conception, may pass the first test. Something can be known yet also unfamiliar or surprising, so the familiarity conception leaves room for the gap between knowing that and understanding why. A further attraction of the familiarity view is the natural way it accounts for the fact that it is often surprise that prompts a request for explanation. It is often when things do not turn out as we expected that we want to know why. Moreover, even when we ask why about what is already in some sense familiar, the prompt for the question often involves `defamiliarisation': we are brought to see the everyday situation as somehow strange or surprising. The case of the moon already mentioned is a good example of this. Most people do not wonder why the same side of the moon always faces the earth, perhaps because they erroneously suppose that this is simply a consequence of the moon not spinning. Once they are shown that, not only does the phenomenon require that the moon spin, but that its period be precisely the same as the apparently unrelated period of the moon's orbit around the earth, the phenomenon becomes surprising and prompts a why-question.

The familiarity conception does not do as well on our other tests. It is unclear whether it allows for self-evidencing explanations. It is difficult to be sure about this without some specific and articulated account of what it takes to make a phenomenon familiar, but if H must itself be familiar in order to explain surprising E, it is unclear how E could provide an

essential part of one's reason for believing H. It is odd to suppose that the surprising provides essential evidence for the familiar.

The familiarity conception also has difficulty with the why regress. If the conception entails that what is familiar is understood and that only what is familiar can explain, then it does not allow that what is not itself understood can nevertheless explain. But the why regress shows that we must allow for this: H may explain E even if we do not understand why H is the case.

The third view on our whirlwind tour is the **unification** conception of understanding. On this view, we come to understand a phenomenon when we see how it fits together with other phenomena into a unified whole (cf. Friedman 1974, Kitcher 1989). This conception chimes with the ancient idea that to understand the world is to see unity that underlies the apparent diversity of the phenomena. The unification conception allows for both the gap between knowledge and understanding and the legitimacy of self-evidencing explanations without difficulty. We can know that something is the case without yet being able to fit it together appropriately with other things we know, so there can be knowledge without understanding. Self-evidencing explanations are also accounted for, since a piece of a pattern may provide evidence for the pattern as a whole, while the description of the whole pattern places the piece in a unifying framework. The unification view may not do quite so well, however, on the why regress. Presumably a unifying explanation is itself unified, so there seems to be no room for explanations that we do not already understand. But this is not clear. For one might say that to explain a phenomenon is to embed it appropriately into a *wider* pattern. In this case H might suitably embed E, even though we have no wider pattern in which to embed H, and the requirements of the why regress would be satisfied.

Our fourth conception of understanding is that of **necessity.** The necessity view is that explanations somehow show that the phenomenon in question *had* to occur (cf. Glymour 1980). This conception of understanding acknowledges the gap between knowing that and understanding why, since one may know that something did in fact occur without knowing that it had to occur. The view also appears to allow for self-evidencing explanations, since there seems to be no vicious circularity involved in supposing that H shows E to be in some sense necessary while E gives a reason for believing H. It is less clear, however, that the necessity conception passes the why regress test: it fails the test if only what is itself necessary can confer necessity, or if only what is already known to be necessary can be used to show that something else is necessary too.

This leaves us with our fifth and final contestant, the **causal** conception of understanding. On this view, to explain something is to give information about its causes (cf. Lewis 1986; Humphreys 1989; Salmon 1998). The causal conception of understanding sails through our three tests. There is a gap between knowing and understanding, because we can know that something occurred without knowing what caused it to occur. The why regress is benign, because we can know that C caused E without knowing what caused C. Self-evidencing explanations are allowed, because it is possible for C to be a cause of E where knowledge of E is an essential part of one's reason for believing that C is indeed a cause.

The relative merits of the different conceptions of understanding are summarized in the following table:

TEST FEATURES	**CONCEPTIONS OF UNDERSTANDING**				
	Reason	**Familiarity**	**Unification**	**Necessity**	**Causation**
Knowledge Versus	NO	YES	YES	YES	YES

Understanding					
Why Regress	NO	NO	MAYBE	NO	YES
Self-Evidencing Explanation	NO	NO	YES	YES	YES

Because it does so well on our tests, and because so many explanations we give both in science and in everyday life are manifestly causal, the causal conception of understanding is my favourite, and will be my focus for the balance of this essay. But the causal conception is not without its difficulties (though I prefer the term `challenges') and, in the spirit of full disclosure, I mention three of them here. The first is that we have no adequate account of causation; the second is that there are some explanations that seem clearly non-causal; the third is that not all causes are explanatory.

The problem of giving an account of the nature of causation is a hardy philosophical perennial. Most recent work is inspired, positively or negatively, by David Hume's enormously influential discussion (1777, Sec. VII). While many philosophers have offered solutions to the problem of the metaphysics of causation, none is generally accepted. (For a collection of recent work, see Sosa & Tooley (eds.) 1993.) The second difficulty for the causal conception of understanding -- the existence of non-causal explanations -- is instantiated by mathematical and philosophical explanations, which are at least usually not causal. There also appear to be physical explanations that are non-causal. Suppose that a bunch of sticks are thrown into the air with a lot of spin, so that they twirl and tumble as they fall. We freeze the scene as the sticks are in free fall and find that appreciably more of them are near the horizontal than the vertical orientation. Why is this? The reason is that there are more ways for a stick to be near the horizontal than the vertical. To see this, consider a single stick with a

fixed midpoint position. There are many ways this stick could be horizontal (spin it around the horizontal plane), but only two ways it could be vertical (up or down). This asymmetry remains for positions near horizontal and vertical, as you can see if you think about the full shell traced out by the stick as it takes all possible orientations. This is a beautiful explanation for the physical distribution of the sticks, but what is doing the explaining are broadly geometrical facts that cannot be causes.

The third and final difficulty for the causal conception is that not all causes are explanatory. Behind every event lies a long and dense causal history, most of which will not explain the event in a given context. When I ask my students why they have failed to hand in their supervision essays on time, I am unimpressed if they respond, `Well, you know about the Big Bang…'.

Nevertheless, I remain a fan of the causal conception of understanding. It is true that we have no adequate metaphysical understanding of causation, but as the why regress teaches us, this does not rule out the use of causal notions to illuminate other things. Nor in my view do we have a better grip on the central notions of any of the other four conceptions of understanding I have canvassed. As for the existence of non-causal explanations, this does show that that causation cannot be the entire story of explanation. As I remarked above, the various conceptions of understanding are not mutually exclusive, so one can opt for more than one. Of the remaining four, the unification conception also did well on our tests, so this is another promising place to look; I also have some sympathy for the necessity conception. It seems clear, however, that very many of the explanations we give cite causes, and that in these cases what is said is explanatory precisely because what is cited is causal information. That leaves us with the difficulty that not all causes are explanatory. This really is in my view more a challenge than a difficulty, and one that we can go some way towards meeting. By giving a

finer grained account of the context in which explanations are requested and of the why questions asked, we can give a causal account of explanation that itself explains why some causes are explanatory and others not. (For further discussion of recent work on explanation and understanding, see Salmon 1989 and Ruben (ed.) 1993.)

Why Do Causes Explain?

As we have seen, the test features of understanding support the causal conception. The gap between knowledge and understanding shows that the goods that explanations provide is more than the good provided by knowledge of the phenomenon to be explained. The why regress shows that the good of understanding is not like a substance that gets transferred from explanation to phenomenon explained, since H can provide an understanding of E even though we do not understand why H itself is the case. Self-evidencing explanations show that understanding does not involve providing a reason for belief. The causal conception respects these facts about understanding, and without portraying understanding as some mysterious form of super-knowledge, since although understanding E is more than knowledge of E, it need be no more than knowledge of the causes of E. Knowledge of causes is a primary good that many explanations provide.

In terms of philosophical explanations, the question we have been asking may be of the form `Why is this a good explanation?', and the answer is `It gives a cause, and causes explain'. This may be a good answer; but it is tempting to take another step up the why regress. Taking that step is to ask why causes explain. But does this question make sense? Or is it like asking why bachelors are unmarried? I think the question why causes explain does make sense, but it is difficult to articulate it in a way that makes this clear, and it is even more difficult to answer the question. I will struggle a bit with both these projects now.

In asking why causes explain, we are continuing our inquiry into the goods of explanation, but the question here does not simply concern the utility of causal knowledge. That question would be too easy. Knowledge of causes is useful for all sorts of reasons; but so is knowledge of effects. Yet while causes explain their effects, effects do not explain their causes. The recession of the galaxy explains why its light is red shifted, but the red shift does not explain why the galaxy is receding, even though the red shift may provide essential evidence of the recession. Part at least of the question I have in mind can be formulated contrastively. Why do causes rather than effects explain? Why don't effects explain their causes, given that causes explain their effects? These are more specific questions than the general question of why causes explain, but they are more than general enough for our purposes.

One may still feel that the question is silly. The reason causes explain and effects do not is simply that `explanation' is a word we apply to causes and not to effects. But this does not do justice to the question. Our explanatory concepts and practices play an enormous role in our cognitive economy, and one wants to know why this is the case. What is the point of this practice? This is just another way of asking about the goods of explanation, and to ask why we privilege causes over effects is a way of getting at part of this question.

Having made a pitch for the question of why causes explain rather than effects, I move briskly from the frying pan into the fire, because the question is very difficult to answer. In particular, it is difficult to avoid a more or less well hidden dormative virtue explanation, along the lines of, `causes explain because they, unlike effects, have the power to confer understanding'. Can we do any better than this? It is not clear. It is certainly not obvious that a thing's effects are any less important, useful or interesting than its causes. And there is a clear sense in which finding out about a thing's effects increases our understanding of that

thing. Indeed one might argue that P's effects typically tell one more about P than do its causes. For effects often give information about P's properties in a way that causes do not. This is so because properties are at least often dispositional, and dispositions are characterised by their effects and not by their causes. Thus to say that arsenic is poisonous is to say roughly that if you eat it you will die. Thus the effects not only lead us back to the properties, but they are constitutive of at least some of them. In the conditional `If you eat it, then you will die' there is both a cause and an effect, but they bear an asymmetrical relation to the corresponding property of being poisonous. Causing death is constitutive of the property of being poisonous, but eating arsenic, though a cause of death, is not constitutive of being poisonous. Nor do the causes of the arsenic or of its presence in a particular place appear to be constitutive of aresenic's properties.

A natural thought is that what is special about the causes of P is that they, unlike P's effects, create or bring about P. Can this be the key to the explanatory asymmetry between causes and effects? One worry is that this may be one of those dormative virtues stories, or worse. Why do causes explain effects? Because causes bring about effects. The worry is that `bring about' is just another expression for `cause', so all that has really been said is that causes explain because they are causes. One response would be to insist on a strong reading of `bring about', a reading that would rule out a Humean account of causation, which takes causation to be no more than constant conjunction. Of course, Humeans may not like this, but they have the option of an error theory of explanation, according to which we never really explain why things happen, though the source of the illusion can be given, much as Hume himself had an error theory of necessary connection, according to which objects in the world are only conjoined, never connected, but the source of our mistaken idea of connection can be given (1777, Sec. VII). Such an error theory of explanation, would treat understanding as a

kind of myth, since it depends on a notion of causation that is metaphysically untenable. This would still be to allow that our notion of explanation and understanding, however misguided, depends on the idea of things being created by their causes. I would find such eliminativism about understanding unpalatable; but not being a committed Humean on matters causal, this line of argument does not overly concern me. Nevertheless, the thought that explanation depends on powerful metaphysical `glue' linking E's cause to E strikes me as problematic for two other reasons. First, as one's account of causation strengthens the link between E's causes and E, it will do likewise for the connection between E and E's effects, so it is not clear that this appeal to a strong connection between cause and effect helps to explain the explanatory asymmetry that concerns us. Secondly, we often explain by appeal to causes that are not strongly connected to what they cause. This is well illustrated by explanatory causes that are omissions. A good answer to the question of why I am eating my campfire meal with a stick is that I forgot to pack my spoon, yet there seems no especially strong metaphysical link between the absence of the spoon and the use of the stick. Of course one may argue that explanations by omissions or negative causes are always oblique references to a positive causal scenario in which the process is strongly creative, but this strikes me as forced.

A closely related but I think better answer to our question of why causes rather than effects explain, though not without difficulties of its own, attributes the special explanatory power of causes to the link between causing and `making a difference'. The idea is that causes explain because causes make the difference between the phenomenon occurring and its not occurring. This is connected to the idea of control, since we control effects through causes that make a difference, causes without which the effect would not occur. P's causes are handles which could in principle have been used to prevent P occurring in a way that P's effects could not. Of course control is not always an option. The galaxy's recession causes

and explains its red shift even though we are in no position to change its motion; but the speed of recession is nevertheless a cause that made the difference between that amount of redshift and another. My suggestion is that this may partially explain why causes rather than effects yield understanding, since causes often make a difference in this sense while effects never do. Information about causes provide a special kind of intellectual handle on phenomena because the causes may make a difference and may themselves provide a special kind of physical handle on those phenomena.

I am far from confident that this difference view is correct, but I have four considerations that may count its favour. First, given the obvious and enormous importance to us of knowledge of practical handles on phenomena, and the close link between control and making a difference, the difference view makes sense of our obsession about explanation. With all our leisure time, this interest has gone far beyond our practical concerns, but this overshooting is not particularly surprising. For one thing, given the difficulty of predicting which handles we will be able in time to pull, a broad strategy makes sense; for another we know that activities or traits originally caused by practical considerations may run way beyond the reasons for which they were originally selected, rather as an inclination to save potentially useful objects may lead to philately.

A second attraction of the difference view is that it may account for our ambivalence about the explanatory use of certain causes. For not all causes do make a difference. The obvious situation where they do not is one of overdetermination. A good ecological example is an environment with foxes and rabbits (Garfinkel 1981, 53-56). We ask why a rabbit is killed; we may answer by giving the location of the guilty fox shortly before the deed, or we may cite the high fox population. Both are causes, but the details of the guilty fox's behaviour does not explain well because, given the high fox population, had that fox not killed the

rabbit, another fox probably would have. Had the fox population been substantially lower, by contrast, the rabbit probably would have survived. The cause that made the difference is the cause that explains. This is some evidence for the difference view, though the situation is not entirely clearcut, since I think we often do judge the actual cause to have some explanatory power even then another cause would have done the job had the first one been absent. One possibility is that although a cause that made the difference is required (or strongly preferred) for explaining why, it is not required for explaining *how*.

The third consideration I adduce in favour of the difference view concerns contrastive explanation and brings out another way in which causes can fail to make the relevant difference and so fail to provide good explanations. Many of the why questions we ask are contrastive. They have the form `Why P *rather than* Q', rather than simply `Why P', though the contrast often remains implicit, because it is obvious in the context in which the question is posed. Moreover, what counts as an explanatory cause depends not just on fact P but also on the foil Q. Thus the increase in temperature might be a good explanation of why the mercury in a thermometer rose rather than fell, but not a good explanation of why it rose rather than breaking the glass. We have already noted that not all of P's causes explain P in a given context; what we see now is that the foil in a contrastive question partially determines which causes are explanatory and which are not. And lo and behold, a good explanation requires a cause that made the difference between the fact and foil (Lipton 1993). Thus the fact that Smith had syphilis may explain why he rather than Jones contracted paresis (a form of partial paralysis), if Jones did not have syphilis; but it will not explain why Smith rather than Doe contracted paresis, if Doe also had syphilis. Contrastive explanations bring out the way in which what makes a difference between the P's occurring or not depends on what we mean by P not occurring, on our choice of foil. In so doing, it seems also to support the idea that the

reason (some) causes explain is that they provide information about what made the salient difference between the occurrence and non-occurrence of the effect of interest.

A final consideration that may support the difference view brings out a perplexing feature of explanation that I have not yet mentioned. This is the opacity of explanation, and it gives yet another way in which a cause (or a causal description) may fail to explain. For whether or not a cause explains depends on how it is described. This is clear, since one way of describing any cause of P is `a cause of P', yet the question `Why did P occur?' is not illuminatingly answered by `P occurred because of its causes'. To take a different example, suppose that the decayed insulation in the high-voltage lines running between the walls caused the fire in the department and is the event mentioned on page 17 of the accident report. If someone asks why the fire occurred, it is unhelpful to say `Because of the event reported on page 17 of the accident report'. That oblique description does refer to a cause of the fire, but the description is not in itself explanatory (cf. Ruben 1990, 162-164).

It is not at all easy to say how we draw the demarcation between explanatory and unexplanatory descriptions of causes, but the idea of making a difference may help here too. The thought is that explanatory descriptions are those where changing the features described would make a difference. It is explanatory to say that the fire in the department occurred because of decayed insulation; it is not explanatory to say that the fire occurred because of the cause mentioned on page 17 of the accident report, however helpful that information may be in finding the explanation. Perhaps this is because, had the insulation not decayed, the fire would not have occurred, whereas it still would have occurred even if its causes where not mentioned in the report. In explanation we want a cause that makes the difference described in a way that tells us in virtue of what the difference is made.

An Instrumental Good of Explanation

Having considered causal knowledge as one good that explanations deliver, and also the question of why causes should explain when effects do not, I end by flagging a quite different sort of good that explanations provide. In a word, this good is inference. This is an instrumental good, not part of understanding, but an example of how our explanatory practices are tools for the acquisition of other valuable things, in this case true beliefs. This is the idea behind Inference to the Best Explanation, an idea I discuss in more detail in `Is Explanation a Guide to Inference?', which appears later in this volume. As I there observe, the model of Inference to the Best Explanation is designed to give a partial account of many inductive inferences, both in science and in ordinary life. Its governing idea is that explanatory considerations are a guide to inference, that scientists infer from the available evidence to the hypothesis which would, if correct, best explain that evidence. Many inferences are naturally described in this way. Darwin inferred the hypothesis of natural selection because, although it was not entailed by his biological evidence, natural selection would provide the best explanation of that evidence. To recycle my astronomical example, when an astronomer infers that a galaxy is receding from the earth with a specified velocity, she does this because the recession would be the best explanation of the observed red-shift of the galaxy's spectrum. When a detective infers that it was Moriarty who committed the crime, he does so because this hypothesis would best explain the fingerprints, blood stains and other forensic evidence. Sherlock Holmes to the contrary, this is not a matter of deduction. The evidence will not entail that Moriarty is to blame, since it always remains possible that someone else was the perpetrator. Nevertheless, Holmes is right to make his inference, since Moriarty's guilt would provide a better explanation of the evidence than would anyone else's (cf. Lipton 1991).

Inference to the Best Explanation can be seen as an extension of one of the three test criteria that I used above to evaluate different notions of understanding. This is the prevalence of self-evidencing explanations, where the phenomenon that is explained in turn provides an essential part of the reason for believing the explanation is correct. According to Inference to the Best Explanation, this is a common situation in science: hypotheses are supported by the very observations they are supposed to explain. Moreover, Inference to the Best Explanation takes the idea of self-evidencing explanations one step further. It is not just that the observations support the hypothesis that explains them; it is precisely because that hypothesis would explain the observations that they support it.

Inference to the Best Explanation thus partially inverts an otherwise natural view of the relationship between inference and explanation. According to that natural view, inference is prior to explanation. First the scientist must decide which hypotheses to accept; then, when called upon to explain some observation, she will draw from her pool of accepted hypotheses. According to Inference to the Best Explanation, by contrast, it is only by asking how well various hypotheses would, if correct, explain the available evidence that she can determine which hypotheses merit acceptance. In this sense, Inference to the Best Explanation has it that explanation is prior to inference, and it is for this reason that inference can be a good that explanations deliver. This view complements the causal view of explanation nicely. Taken together, we have the idea that the construction and evaluation of competing explanations is one important route to the discovery of causes. If our explanatory practices give us this sort of information, it is unsurprising that they play such a large role in our cognitive economy.

Conclusion

By asking about the goods of explanation, I have been seeking a kind of explanation of

explanation. Philosophical explanations are perhaps particularly prone to the flaw of dormative virtues, where opium puts people to sleep because of its dormative powers. To say that we value explanations because they provide understanding is this sort of an inauspicious beginning. In the absence of an independent account of understanding, it gives us little more than the observation that we value explanations because of their explanatory power. This is the reason I began by considering different accounts of what understanding amounts to. Having settled on the causal view, I then considered the vexing question of why we should find causes explanatory and in particular why causes explain while effects do not, suggesting that only causes can make the relevant difference between the occurrence and non-occurrence of the thing we want explained. I then briefly suggested another explanatory good of a quite different order: not a type of understanding, but what understanding, especially causal understanding, is good for. It is good for causal inference.

This sort of explanations of explanation I have sketched avoid dormative virtues, since the notions of causation, making a difference and inference have the requisite independence from the notion of explanation itself. But what kind of explanations have I provided? Surprisingly perhaps, at least one of them is itself causal. I have suggested that one of the functions of explanation is inference. Functions are effects, but I go along with the view that function explanations are nevertheless causal, not `effectal'. In the biological case, to explain the presence of a trait functionally is sometimes to use an effect as an oblique way of giving information about the evolutionary etiology of the trait itself. Thus to say that the function of a polar bear's white fur is camouflage is to explain the presence of fur of that colour in terms of a causal history of evolution in which the possession of such fur by earlier bears or their progenitors conferred a selective advantage and so caused there to be later bears with the same trait. To say that inference is a function of explanation may likewise be to provide a

broadly causal explanation of the prevalence and persistence of our explanatory practices. Asking why things are as we find them to be provides us with an important way of discovering causes, and the fact that explanatory practices have this power is one of the reasons those practices have such a grip on us. It is curiously satisfying that we may thus give a causal explanation of causal explanation.

Acknowledgements: I am grateful to the Giora Hon, Sam Rakover and Wesley Salmon for very helpful comments on an earlier draft of this paper.

References

Friedman, Michael, 1974. `Explanation and Scientific Understanding', *The Journal of Philosophy*, **71**, 1-19.

Garfinkel, Alan, 1981. *Forms of Explanation*. New Haven: Yale University Press.

Glymour, Clark, 1980. `Explanations, Tests, Unity and Necessity', *Nous*, **14**, 31-50.

Hempel, Carl, 1965. *Aspects of Scientific Explanation*. New York: Free Press.

Hume, David, 1777. *An Enquiry Concerning Human Understanding*, L.A. Selby-Bigg & P.H. Nidditch (eds.), 1975, Oxford: Oxford University Press.

Humphreys, Paul, 1989. *The Chances of Explanation*. Princeton: Princeton University Press.

Kitcher, Philip, 1989. `Explanatory Unification and the Causal Structure of the World', in Kitcher & Salmon (eds.), 1989, 410-505.

Kitcher, Philip & Salmon, Wesley (eds.), 1989. *Scientific Explanation*, Vol 13, *Minnesota Studies in the Philosophy of Science*. Minneapolis: University of Minnesota Press.

Lewis, David, 1986. `Causal Explanation', in his *Philosophical Papers*, Vol. II, New York: Oxford University Press, 214-240.

Lipton, Peter, 1991. *Inference to the Best Explanation.* London: Routledge. Expanded 2nd edition 2004.

Lipton, Peter, 1993. `Contrastive Explanation', in Ruben (ed.), 1993.

Ruben, David-Hillel, 1990. *Explaining Explanation*. London: Routledge.

Ruben, David-Hillel (ed.), 1993. *Explanation.* Oxford: Oxford University Press.

Salmon, Welsey, 1989. *Four Decades of Scientific Explanation*. Minneapolis: University of Minnesota Press.

Salmon, Wesley, 1998. *Causality and Explanation.* New York: Oxford University Press.

Sosa, Ernest & Tooley, Michael (eds.), 1993. *Causation.* Oxford: Oxford University Press.