

# The Fine Structure of Inference to the Best Explanation

STATHIS PSILLOS  
*University of Athens*

## The Search for the Holy Grail

Traditionally, philosophers have focused mostly on the logical template of inference. The paradigm-case has been deductive inference, which is topic-neutral and context-insensitive. The study of deductive rules has engendered the search for the Holy Grail: syntactic and topic-neutral accounts of all *prima facie* reasonable inferential rules. The search has hoped to find rules that are transparent and algorithmic, and whose following will just be a matter of grasping their logical form. Part of the search for the Holy Grail has been to show that the so-called scientific method can be formalised in a topic-neutral way. We are all familiar with Carnap's inductive logic, or Popper's deductivism or the Bayesian account of scientific method.

There is no Holy Grail to be found. There are many reasons for this pessimistic conclusion. *First*, it is questionable that deductive rules are rules of *inference*. *Second*, deductive logic is about updating one's belief corpus in a consistent manner and not about what one has reasons to believe *simpliciter*. *Third*, as Duhem was the first to note, the so-called scientific method is far from algorithmic and logically transparent. *Fourth*, all attempts to advance coherent and counterexample-free abstract accounts of scientific method have failed. All competing accounts seem to capture some facets of scientific method, but none can tell the full story. *Fifth*, though the new Dogma, Bayesianism, aims to offer a logical template (Bayes's theorem plus conditionalisation on the evidence) that captures the essential features of non-deductive inference, it is betrayed by its topic-neutrality. It supplements deductive coherence with the *logical* demand for probabilistic coherence among one's degrees of belief. But this extended sense of coherence is (almost) silent on what an agent must infer or believe.

The root of the problem with the search for the Holy Grail is that there is a type of inference broadly called 'defeasible', or 'ampliative'

or ‘non-monotonic’, with the following characteristics. *One*, it is psychologically real and compelling. *Two*, it is reasonable, that is it can offer at least *prima facie* entitlement to form new beliefs and justification for having them. *Three*, it is sensitive to the context, background information, cognitive aims and values etc. *Four*, it works, (it produces truths or likely truths), only if the external circumstances are right (if the world co-operates). *Five*, it offers warrant that can be revoked in the presence of new information and hence requires taking into account considerations external to the inference itself, e.g., whether the evidence is total, whether there are defeaters etc. The key point is that this type of inference does not admit of a simple abstract-logical form. Nor does it admit of an analysis suitable for deductive logic. Its details are too messy to admit of a neat formal characterisation, but its strength is precisely in those details. Unlike deductive reasoning and Bayesian updating (which are both non-ampliative), defeasible inference has the *prima facie* advantage that it takes us to beliefs with new content. It does not start with the assumption that reasoning (and inference) is about updating an already existing (learned at our mother’s knee?) belief corpus in light of new information. It plays an active role in creating and maintaining this belief-corpus.

### **Inference to the Best Explanation**

If anything, Inference to the Best Explanation (henceforth, IBE) is an instance of this general type of defeasible reasoning. It should then be investigated as such. Any account of it should respect that it is meant to be a candidate for an ampliative and context-sensitive defeasible rule of inference. In his excellent (2004), Lipton stresses that the guiding idea behind IBE is that explanatory considerations guide inference. That is, a hypothesis is accepted on the basis of a judgement that it bests explains the available evidence.

Lipton rightly emphasises the difference between loveliness and likeliness. But he feels the pressure to show that explanatory loveliness is conducive to likeliness. How can the fact that a hypothesis is the best explanation of the evidence make it likely to be true? Here, I think, the pressure is largely due to a feeling that all of us who defend IBE have had, viz., that we have to offer an abstract (topic-neutral) template for it. Suppose we reason thus: IBE is the mode of inference that proceeds as follows.

D is a collection of data (facts, observations).

H explains D. (H would, if true, explain D.)

No other hypothesis can explain D as well as H does.

---

Therefore, H is (probably) true.

Then, there is certainly a pressure to analyse the inferential relation in terms of the likeliness of the conclusion. How for instance are we to understand the qualification ‘no other hypothesis’? Are we to include the catch-all hypothesis? And if yes, how are we to determine its own probability? Or, how are we to understand the relation ‘H explains D’? And shouldn’t H be a satisfactory explanation independently of its possible competitors?

There is a sense in which these are genuine questions (cf. my 2002). Indeed, if we think that IBE has an abstract logical structure we *should* address them (and Lipton does address them in detail). But we could approach the issue from another route. We could examine specific cases of defeasible reasoning in which explanatory considerations play a key role. These cases reveal that an IBE-type of reasoning has a fine structure that is shaped, by and large, by the context. Explanations are, by and large, detailed stories. The background knowledge (or, beliefs) ranks the competitors. Other background assumptions determine the part of the logical space that we look for competitors. The relevant virtues or epistemic values are fixed, etc. Given this rich context, one can conclude, for instance, that the double-helix model is the best explanation of the relevant evidence, or that the recession of the distant stars is the best explanation of the red-shift. What *is* important in cases such as these is that the *context* a) can show how explanatory considerations guide inference; b) can settle what is the relevant explanatory relation; c) can (in most typical cases) determine the ranking of the rival explanations; d) can settle what assumptions must be in place for the best explanation to be acceptable; e) can settle what to watch out for (or search) before the best explanation is accepted (e.g., the absence of certain defeaters). These contextual factors can link loveliness and likeliness nicely, because they do not try to forge an abstract connection between them; rather the connection stands or falls together with the richness and specificity of the relevant information available.

It is after we are clear on the rich fine structure of specific inferences that we can start extracting (or abstracting) a common inferential pattern of defeasible reasoning based on explanatory considerations. We can then write down a schema such as the above. We can talk about the features that a candidate should have in order to be the best explanation. We can, for instance, talk in terms of two sorts of filter (as Lipton rightly notes) through which the inference proceeds, viz., one that selects a relatively small number of potential explanations as

plausible, and another that selects the best among them as the actual explanation, where both filters operate with explanatory considerations. In offering this abstract template we distil what is common in the many concrete cases. The key idea is that IBE rests on an explanatory-quality test. It is the *explanatory quality* of a hypothesis, on its own but also taken in comparison to others, which contributes essentially to the warrant for its acceptability. The advantage of proceeding the way suggested is that there is no longer (great) pressure to answer in the abstract the question about the link between loveliness and likeliness. It should be implicit in the abstraction that, if anything, the context of application will settle the question: how likely is the best explanation?

Seen in this light, IBE should be considered as an inferential *genus*. The several species of the genus IBE are distinguished, among other things, by plugging assorted conceptions of explanation in the reasoning schema that constitutes the genus. For instance, if the relevant notion of explanation is causal, IBE becomes an inference to the best causal explanation. Or, if the relevant notion of explanation is subsumption under laws, IBE becomes a kind of inference to the best nomological explanation, and so forth. This might be taken to be an *ad hoc* attitude to the apparently pressing issue of the model of explanation that features in IBE. But it is not. Many things can be explanations and the explanatory relation is not fixed and immutable. Attempts to cast the concept of explanation under an abstract model are parts of the search for the Holy Grail. For one, there is no extant model of explanation that covers all aspects of the explanatory relation. It is not implausible to think that ‘explanation’ is a cluster-concept, which brings together many distinct but resembling explanatory relations. For another, as Lipton rightly stresses, IBE should not be seen as issuing commitment to a single explanation. Rather, the commitment is to one (the best) among *competing* explanations. An extension of Lipton’s thought might well go as follows: an explanandum might admit of different *types* of explanation (causal, nomological, mechanistic, unificatory etc.). There is no reason to take one of them as being the right type of explanation. In many cases, different types of explanation will be compatible with each other. Which of them will be preferred will depend on interests and on the details of the case at hand.

Lipton favours causal explanation (see chapter 3). As he is aware, there are problems that the causal account of explanation faces, the most obvious being that there are non-causal explanations (for instance, the explanation of laws in terms of other laws). Lipton carries on regardless and does a fine job in analysing causal explanation in terms of contrasts: why P *rather than* Q? I will not pause to examine his contrastive account, but I cannot help noting that, for all its

advantages, the contrastive model fails to capture the fact that in many typical cases of scientific explanation the task is to answer a ‘straight’ why-question: *why P?* If we view an explanation as something *in* the world (a cause, or a law, or a statistical pattern, or a causal history) in virtue of which the explanandum is accounted for, then we ultimately demand an answer to a ‘straight’ why-question; the contrastive pattern is helpful only as a means to find out the answer to this question. Indeed, it seems that the role of contrasts is to highlight that IBE is, at bottom, an eliminative method: that an explanation emerges as the best by a process of eliminating alternative potential explanations. Lipton (2004, 90) comes close to acknowledging this. But then it seems that what matters is the very idea of the elimination of alternatives and not how *exactly* we go about doing it, that is by Mill’s methods, or by clinical trials, or by deriving further predictions, by looking for defeaters etc. My diagnosis is that Lipton is carried away by the similarities between contrastive causal explanation (which invites the thought that contrasts can be explained by different *causal* hypotheses—e.g. *P rather than Q* because *F* was present in *P*’s history but *F* was absent in *Q*’s history) and the eliminative dimension in IBE (which invites the thought that accepting *any* explanation—be it causal or not—rests on eliminating alternatives). Contrastive causal inference is just one type of IBE, seen as an eliminative inferential genus. In the end, what matters for the contrastive inference is not that the explanation is causal but that it is not eliminated.

To press this point a bit more, what *is* important, it seems to me, is not the type of explanation preferred but the *role* that explaining has in our inferential practices and our overall conceptual economy. Whatever the formal details of an act of explaining, it should incorporate the explanandum into the rest of the reasoner’s background knowledge by providing some link (even by *breaking* a link) between the explanandum and other hypotheses that are part of the reasoner’s background knowledge. Accordingly, explanation has a coherence-enhancing role. Explanatory coherence is a cognitive virtue because it is a prime way to confer justification on a belief or a corpus of beliefs.

### Bayesianism

My real disagreement with Lipton concerns his irenic stance towards Bayesianism. Unlike him, I would not go out my way to argue that explanationists and Bayesians should be friends. I have argued elsewhere (2004) why I think so. My chief objections to the irenic approach are three. *First*, on a strict Bayesian approach, we can never detach the probability of the conclusion of a probabilistic argument, no matter

how high this probability might be. So, strictly speaking, we are never licensed to *accept* a hypothesis on the basis of the evidence. (Unless, of course, we institute well-known and problematic rules of acceptance.) All we are entitled to do is a) detach a conclusion about a probability, viz., assert that the posterior probability of a hypothesis is thus-and-so; and b) keep updating the posterior probability following Bayesian conditionalisation on fresh evidence. But IBE is a *rule of acceptance*: it authorises the *acceptance* of a hypothesis, on the basis that it is the best explanation of the evidence. Why worry about acceptance? Lipton (2004, 63) does not. Bent on a reconciliation between IBE and Bayesianism, he argues that “explanatory considerations provide a central heuristic we use to follow the process of conditionalisation” (2004, 107), but that the “mechanics of belief revision” follows the Bayesian route. So, in the end, Lipton talks about degrees of confirmation (comparisons between posterior and prior probabilities) and adds that he writes about IBE “as if” it were “a mechanism of acceptance” (cf. 2004, 63). I think this is a big concession to Bayesianism. Defeasible reasoning is *not* (primarily) about degrees of confirmation. Clearly, this is true descriptively. But it should be true normatively as well. Defeasible reasoning is about what we are entitled to believe in light of reasons (and evidence) that are *not* deductive. (Incremental) confirmation is neither sufficient nor necessary for reasonable belief. It is not sufficient because though a piece of evidence might increase the probability of a hypothesis *H*, it does not *ipso facto* make belief in *H* reasonable. And it is not necessary because hanging on to a hypothesis *H* even if it is disconfirmed by a piece of evidence might well be reasonable if there is no alternative to replace it.

*Second*, the Bayesian approach is not ampliative: it counsels a purely logical update of degrees of belief. IBE, on the other hand, is an *ampliative* mode of reasoning. It is supposed to deliver informative hypotheses and theories, whose content exceeds the observations, data, experimental results etc. that prompted them. This content-increasing aspect of IBE is indispensable, if science is seen, at least *prima facie*, as an activity that purports to extend our knowledge (and our understanding) beyond what is observed. Bayesian reasoning does not have the resources to be ampliative. All is concerned with is maintaining synchronic consistency in a belief corpus and (for some Bayesians) achieving diachronic consistency too. It may be argued that Bayesian reasoning allows for ampliation, since this can be expressed in the choice of hypotheses over which prior probabilities are distributed. In other words, one can assign prior probabilities to ampliative hypotheses and then use Bayesian kinematics to specify their posterior probabilities. But even if this is granted, the attempt to place IBE in a

Bayesian framework *that* way concedes that explanatory considerations operate *only* in the context of discovery, as a means to generate plausible ampliative hypotheses and distil the best among them. Then, the best explanation is taken over to the context of justification, by being embedded in the Bayesian framework of confirmation, which determines its credibility. I think the friends of IBE should take it to be both ampliative *and* warrant-conferring at the same time.

*Third*, though there is no doubt that the Bayesian approach can accommodate explanationist concerns within itself, this should be neither surprising nor particularly enlightening. The Bayesian framework (especially the machinery of subjective prior probabilities) is flexible enough to accommodate almost *everything*. For instance, as Lipton (2004, 115) notes, it may be argued that explanatory considerations play some role in fixing prior probabilities. But we should be careful here. It is one thing to say that priors are informed by explanatory considerations, another to say that they *should* be so informed. No-one would doubt the former, but subjective Bayesianism is bound to deny the latter. So, we come to the crux. There are two ways to place IBE within a Bayesian framework. The *first* pays only lip service to explanatory considerations. All the work in degree-of-belief updating is done by the usual Bayesian techniques. It may be admitted that the original assignment of prior probabilities might be influenced by explanatory considerations, but the latter are no less idiosyncratic (from the point of view of the subjective Bayesian) than specifying the priors by, say, consulting a soothsayer. If we think this way, IBE loses much of its excitement. It just highlights the *permission* to use explanatory considerations in fixing an initial distribution of prior probabilities. (I am not suggesting that Lipton's conciliatory account pays only lip-service to explanatory considerations. After all, he insists that explanatory considerations are doing real psychological work in the way we do what the Bayesians try to describe. But there is a sense in which, even in Lipton's rich account, the inferential engine is primarily driven by the Bayesian machinery and only secondarily by the explanatory considerations.) The *other* way to place IBE within a Bayesian framework is to take explanatory considerations to be a *normative* constraint on the specification of priors. This would be an exciting way to supplement Bayesianism with explanationism. It would capture the idea that explanatory considerations should be a constraint on *rational* belief. But it's obvious that *this* way amounts to a radical modification of the current Bayesian orthodoxy.

In the Bayesian framework, there are also the likelihoods to play with. One could argue that explanatory considerations can help determine the likelihoods of competing hypotheses. But there is no easy way

to capture the claim that the best explanation is the one with the highest likelihood because this could easily lead to the base-rate fallacy (cf. my 2004). Lipton considers two more ways in which the explanationist can help the Bayesian: the determination of the relevant evidence and the guidance for construction of fertile hypotheses. Both, however, would take us outside the strict Bayesian framework. For Bayesians, it is simply *not* an issue what kinds of pieces of evidence are relevant to a hypothesis. The only concern is this: *given* the evidence, how are the probabilities already assigned to hypotheses re-distributed? (Here Lipton rightly suggests that a piece of evidence becomes relevant to a hypothesis by seeing that the hypothesis would explain it.) When it comes to preference for fertile (that is, strong in content) hypotheses, it should be obvious that we need to look to more than probabilistic considerations to account for it.

I conclude with a dilemma: *either* accommodate (relatively easily) IBE within Bayesianism, but lose the excitement and most of the putative force of IBE *or* endorse an interesting version of IBE but radically modify Bayesianism. My feeling is that Lipton (rightly) favours the second horn. But if so, IBE should be a lot more than an “ explanationist crutch” (Lipton 2004, 134) to Bayesian reasoning. Lipton, to be sure, uses the “ explanationist crutch” to indicate a weakness in Bayesianism and *not* in IBE (Bayes’s theorem needs an explanationist crutch to carry some of its weight), but the crutch metaphor wavers between highlighting a permanent incapacity of Bayesian reasoning and highlighting a temporary aid that should be dispensed with in the proper account of our inferential practices.

### References

- Lipton, Peter (2004) *Inference to the Best Explanation* (Second revised edition), London: Routledge.
- Psillos, Stathis (2002) ‘Simply the Best: A Case for Abduction’ in A. C. Kakas and F. Sadri (eds.) *Computational Logic: From Logic Programming into the Future*, LNAI 2408, Berlin-Heidelberg: Springer-Verlag, pp. 605-625.
- Psillos, Stathis (2004) ‘Inference to the Best Explanation and Bayesianism’ in F. Stadler (ed.) *Induction and Deduction in the Sciences*, Kluwer, pp. 83-91.