

Explanatory reasoning in the material theory of induction

One of the many exciting ideas in John Norton's book is his account of inference to the best explanation (IBE). A natural approach to understanding IBE is to clarify explanatoriness, and then determine how we might be able to use comparative explanatory success for inference. However, Norton suggests a different view: IBE is warranted when a theory adequately accommodates our evidence, whereas alternative explanations fail to do so. His view apparently implies that IBE is sophisticated eliminative induction, though Norton does not use this term. Thus, explanatoriness plays no role in his account (248). I shall argue that this feature of his account offers a surprising possibility: we can use our understanding of IBE as an element for analysing explanation, without circularity.

Yet there are famous arguments aiming to prove that a theory's accommodation of evidence is insufficient for explanation. One might thus suspect that Norton's view is incomplete as a theory of inference to the best *explanation*. However, I shall argue that the classic arguments do not prove the necessity of adding a notion of explanatoriness to his account of IBE.

First, I briefly summarise his account of IBE. Second, I illustrate one of its advantages. Third, I argue that an additional notion of explanatoriness is unnecessary. While Norton's account fits well with his Material Theory of Induction (MTI), it is compatible with many theories of induction. Therefore, I recommend reading further even if you are critical of the MTI.

Norton's Account of IBE

The account is developed in two chapters of Norton's book. In Chapter 8, he describes the general features of IBE. In Chapter 9, he details some case studies that inspired his generalisations. As he notes, one positive feature of these case studies is that they feature plausible alternatives to the inferred theories. In contrast, many alleged examples of IBE in the literature lack serious competition between rival theories, and thereby fail to reveal the subtleties involved in real scientists' reasoning.

Norton argues that IBE is not a distinct inference form. Instead, he observes that good IBEs have two general features:

(1) An inferred theory T accommodates a challenging body of evidence, relative to total relevant evidence E . This accommodation can either be that the data is expected given $(T \wedge E)$ or that merely its *possibility* was expectable given $(T \wedge E)$. Often, T has "evidential debts", i.e. ancillary and currently unestablished hypotheses that must be added to E for the accommodation. For instance, in *The Origin of Species*, Darwin detailed how his theory of speciation could accommodate a huge number of diverse and surprising facts. However, famously, there were some important evidential debts that were not paid until later discoveries of radioactivity and Mendelian genetics.

(2) There are severe problems with competitors to T . They may fail to accommodate the existing evidence. For example, Darwin knew overwhelming evidence against any diluvial theories—those that tried to reconcile geology and palaeontology with a literal belief in the Biblical Flood story (Herbert 2005, 188). Another possible

problem is that the prospective payment of competitors' evidential debts may be implausible. Biblical literalism could accommodate much of Darwin's evidence, but only via ancillary hypotheses that he could have reasonably predicted would never be established, such as explaining the speciation since the putative time of Noah's Ark by an extraordinarily rapid rate of evolution in places like New Zealand and the Galapagos islands. Norton does not specify the requirement that the competitors' payment of their debts is implausible, whereas *T*'s payment is plausible, but something like this requirement is necessary to avoid arbitrariness. For example, Norton notes how Darwin's theory of speciation and the rival theory of an independent creation of each species both had some evidential debts (278–282).

Norton's account of IBE is not intended as a universal analysis of its necessary and sufficient conditions. It is an inductive extrapolation from some paradigmatic case studies (Norton 2021, 116–117). However, his observations suggest that IBE might be reducible to eliminative induction. By the latter, I do not mean some purely formal argument schema, but rather when our evidence selects, tentatively, some theory over its known competitors, and our evidence is suitably strong. Norton has correctly emphasised that local background knowledge is required for eliminative induction.

Since his theory of IBE involves two parts, (1) the accommodative part and (2) the eliminative part, I shall call his theory the Accommodation-Elimination account of IBE. A full discussion of this account is beyond my scope, but I shall note a potentially advantageous feature: "explanation" does not feature in it. Norton does not require that inferred theories must satisfy some independent definition of explanation. Consequently, we can use the Accommodation-Elimination account to understand scientific explanation, without

circularity. In particular, I suggest that the following may be a necessary condition for *theoretical* explanation:

Candidacy: A theory T explains data D relative to total evidence E and initial conditions I only if (i) we can expect D given $(T \wedge E \wedge I)$ and (ii) T is not eliminated by $(D \wedge E)$. In short, T must be a live candidate for IBE.

What does this requirement give us? It explains the apparent historical relativity of claims of theoretical explanation. For example, it seems intuitive to say (a) phlogiston theory *explained* many phenomena of combustion and rusting, but not (b) phlogiston theory *explains* these phenomena today. Intuition (b) is entailed by a requirement that explanations must be true, but that requirement is incompatible with (a) (see also van Fraassen 1980, 98).

Candidacy allows both. This is closer to how historians of science and scientists themselves talk about explanation (e.g. Deming 2016, 107).

An alternative view would be that theories are only explanatory at some time if they are highly confirmed by the available evidence. This preserves historical relativity, but faces counterexamples like phrenology, which explained some phenomena but was never highly confirmed. In contrast, *Candidacy* would say that phrenology explained some phenomena, although subsequent evidence ended its viability. Thus, *Candidacy* exemplifies how Norton's Accommodation-Elimination account opens up some intriguing possibilities.

Can Accommodationism Accommodate Explanation?

There are some features of Norton's account that need more clarification. When does T accommodate some data D ? Presumably, in conjunction with other premises, some of which may be speculative, T should provide evidence for D . For example, Einstein's general theory of relativity, in conjunction with astronomical facts (initial conditions) explained the anomalous data about Mercury's orbit, because general relativity theory would be evidence for the data, relative to the background astronomy. I mean "evidence" in a strong sense: a theory T is evidence for data D relative to background assumptions A if and only if we should expect D assuming that $(T \wedge A)$ is true, but not if we just assume A . Mere probabilistic relevance of T to D relative to A is insufficient. Unlike philosophers like Carl Hempel (1965), who thought that the support of D by T is reducible to purely formal deductive or probabilistic relations, Norton correctly emphasises that these evidential relations are determined by the local material facts.

If the Accommodation-Elimination account is correct, then accommodating the specific data and not being eliminated by its conjunction with our total evidence is sufficient to be a candidate for a reasonable IBE. I shall call this idea "accommodationism". Yet there are numerous counterexamples that allegedly show that these criteria are neither necessary nor sufficient for being a potential best explanation. Do they prove the necessity of an additional notion of explanatoriness (ontic, pragmatic, or whatever) to identify candidate theories for IBE? Surprisingly, they do not.

Specificity: the example of birth control pills

That all males who take birth control pills do not become pregnant explains why, as a male, Hempel did not become pregnant.

Is accommodationism committed to this claim? The theory cited in *Birth Control-Pills*, “All human males who take birth control pills do not become pregnant,” can be interpreted in multiple ways. Perhaps the most plausible is that it asserts a causal connection between taking birth control pills and non-pregnancy in males. In that case, it is inconsistent with our total evidence, and therefore it cannot be an explanation. Note that this is not the same as requiring that *all* explanatory theories must posit a plausible causal mechanism; the issue is with the implausibility of this particular causal mechanism.

What if we interpret this theory as a non-causal generalisation: as denying the possibility of any males who have taken birth control pills and become pregnant, but not due to link between the pills and the non-pregnancy? With that interpretation, there is a linguistically pragmatic problem with asserting *Birth Control-Pills*. Its assertion suggests that there are human males, who do not take birth control pills, who do become pregnant. This is the *prima facie* reason for the circumscription to a subset of males. Additionally, its assertion may falsely suggest some uncertainty about the possibility of pregnancy in some subsets of males. Given these conversational implications, asserting *Birth Control-Pills* violates two Gricean maxims: the reference to a subset of males rather than males in general is needlessly prolix one could convey more information more briefly by just referring to males and involves a circumscription that is presumably irrelevant in almost any plausible conversation (Grice 1989, 26). I am not assuming the correctness of Grice’s analysis or general approach. I

mention his ideas merely to illustrate how we might think that there are false pragmatic implications of *Birth Control-Pills*. Due to these implications, we naturally recoil from *Birth Control-Pills*. Thus, while *Birth Control Pills* is not false, its assertion would be needlessly misleading in the conversational contexts that we assume it would be asserted. Consequently, *Birth Control-Pills* is no counterexample to accommodationism: it is either eliminated or it is unassertible.

Asymmetry: the example of shadow

Imagine a flagpole whose height is x and shadow's length is y . Suppose that the flagpole is in an otherwise empty warehouse room, with a single light source shining on the flagpole with an angle z . Using L , the laws of light's rectilinear propagation, we can deduce x 's value from these initial conditions and values of the magnitudes y and z , using the formula $x = y \tan(z)$, assuming that all variables exceed 0. Therefore, we can explain the flagpole's height by its shadow.

This type of example is supposed to prove that accommodationism is too broad (Bromberger 1970). However, the counterexample is not true for accommodationism about theoretical explanation in particular, because neither x nor y are theories. This point can be obscured by sliding from explanations by theories to explanations by facts: the usually careful Wesley C. Salmon slides from one to the other in three sentences (Salmon 1989, 47). However, one might also object to this refashioning:

Law+Conditions The aforementioned initial conditions and L explain x (Woodward 2005, 191).

Yet, in addition to being precise about the explanans, we must be precise about the explanandum. If the explanandum is that, at a time t and given these initial conditions, the magnitude of the flagpole's length must be y , then L does seem to be explanatory. L explains why y could not be otherwise *under these conditions*. It explains just a snapshot fact about the flagpole, but that is all there is to explain in the explanandum. On this interpretation, *Law+Conditions* is true, but it is no reason to reject accommodationism.

Alternatively, the explanandum could be the origin of the flagpole's length, i.e. the process by which the flagpole's height became x . In this case, given only the initial conditions provided in the example, L does not accommodate the evidence. In whatever way we detail it, an explanandum of origins will involve temporal processes, but since L consists of laws of coexistence rather than laws of succession, it is clearly unsuitable for the task. On these interpretations of the explanandum, L does not accommodate the evidence, and hence *Law+Conditions* is not a counterexample to accommodationism (for a similar analysis, see Levin and Levin 1977, 294). True, we may say things like "The flagpole explains the shadow but not vice versa," but this is not a claim about theoretical explanation but explaining phenomena by other phenomena.

Statistical relevance: the example of uranium

Given that half-life of uranium U^{238} is about $(4.5)10^9$ years, atomic physics asserts a low probability that a small sample of U^{238} will emit radiation in a brief interval, so atomic physics does not explain any such emission (van Fraassen 1980, 105).

Similar examples stimulated Salmon to try to analyse explanation in terms of probabilistic relevance rather than accommodation (Salmon 1970). Hempel dismissed such counterexamples to his version of accommodationism, alleging that they were like saying that buying a single ticket in a large lottery explains one's winning that lottery (Hempel 1965, 369–370). So there is a clash of intuitions.

Again, we need to be careful in specifying the explanandum. Is it (1) an emission in that particular unlikely time interval? Or (2) that a particular emission in that time interval was *possible*? Modern physics does not predict (1), but it does predict (2). Nor is this latter prediction trivial: predicting how something as strange as radioactive decay can occur is one of the great achievements of modern physics. However, it does not seem to explain why an emission occurred in some particular improbable time period. Anecdotally, some physicists give such ontically indeterministic events as examples of how there are inevitably unexplainable facts according to modern physics. Hempel was right that the relevant theories do not explain the emissions themselves, but there is an explanatory success regarding the *possibility* of the emissions.

Conclusion

In addressing these counterexamples, I have not claimed that accommodationism offers an adequate general account of explanation. Instead, I have argued that the counterexamples do not prove that some additional notion of “explanatoriness” is necessary for identifying candidate explanation for IBE. My focus has been theoretical explanations, because these are the explanations which seem to confirm theories, and which one might think require the addition of a concept of “explanatoriness” in the MTI theory of evidence. I have argued that, in the case of theoretical explanations, Norton is right: no such concept is necessary. There are also some examples that I have not discussed, because they are about explaining phenomena by other phenomena (Scriven 1959, 456; for recent discussion, see Hu 2021, 167).

In general, there are many further issues about explanation and IBE that need exploration from an MTI perspective. For instance, evidential debts are a ripe topic for research. Why were the evidential debts of Einsteinian physics less important than those of Newtonian physics in the 1910s and 1920s? True, these contrasts will intimately involve the relevant local facts, but perhaps there are patterns. Additionally, there are debates about whether the elimination step is ever sufficient to justify inferring a theory. I suspect that Norton’s account of IBE, when expanded and combined with the MTI, can clarify whether this step is rational. I hope that he looks forward to exploring these avenues that are opened by his book as much as I do.

References

Bromberger, Sylvain. 1970. Why-Questions. In Baruch A. Brody, ed., *Readings in the Philosophy of Science*, 66–84. Prentice Hall, Inc.

Deming, David. 2016. *Science and Technology in World History, Volume 4: The Origin of Chemistry, the Principle of Progress, the Enlightenment and the Industrial Revolution*. McFarland & Company, Inc.

Grice, Paul. 1989. *Studies in the Way of Words*. Harvard University Press.

Hempel, Carl. 1965. *Aspects of Scientific Explanation*. Collier-Macmillan.

Herbert, Sandra. 2005. *Charles Darwin, Geologist*. Cornell University Press.

Hu, Xingming. 2021. Hempel on Scientific Understanding. *Studies in History and Philosophy of Science* 88, pp. 164–171.

Levin, Michael E. and Levin, Margarita R. 1977. Flagpoles, Shadows and Deductive Explanation. *Philosophical Studies* 32, pp. 293–299.

Norton, John. D. 2021. Author's Responses. *Studies in History and Philosophy of Science* 85, pp. 114–126.

Salmon, Wesley. C. 1970. Statistical Explanation. In Robert Colodny (ed.), *The Nature and Function of Scientific Theories*. University of Pittsburgh Press.

Scriven, Michael. 1959. Explanation and prediction in evolutionary theory: Satisfactory explanation of the past is possible even when prediction of the future is impossible. *Science* 130, pp. 477–482.

Van Fraassen, Bas C. 1980. *The Scientific Image*. Oxford University Press.

Woodward, James. 2005. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press.