Is there a Bayesian justification of hypothetico-deductive inference?

Samir Okasha*[1] and Karim Thébault[1]

[1] *Department of Philosophy*, University of Bristol, Bristol BS6 5DR, U.K.

Abstract

Many philosophers have claimed that Bayesianism can provide a simple justification for hypothetico-deductive (H-D) inference, long regarded as a cornerstone of the scientific method. Following up a remark of van Fraassen (1985), we analyze a problem for the putative Bayesian justification of H-D inference in the case where what we learn from observation is logically stronger than what our theory implies. Firstly, we demonstrate that in such cases the simple Bayesian justification does not necessarily apply. Secondly, we identify a set of sufficient conditions for the mismatch in logical strength to be justifiably ignored as a "harmless idealization". Thirdly, we argue, based upon scientific examples, that the pattern of H-D inference of which there is a ready Bayesian justification is only rarely the pattern that one actually finds at work in science. Whatever the other virtues of Bayesianism, the idea that it yields a simple justification of a pervasive pattern of scientific inference appears to have been oversold.

*Lead author, email: Samir.Okasha@bristol.ac.uk

Contents

1 Introduction

A venerable tradition in the philosophy of science offers the following schematic description of the scientific method. Scientists propose a theory about the world; they deduce testable empirical predictions from the theory; they then perform observations or do experiments to see if the predictions are true or false; if they are false then the theory is rejected, while if they are true then the theory is confirmed or supported, at least to some extent. This of course is the famous "hypothetico-deductive method" (H-D method), or the "method of hypothesis" as it was called in previous centuries; its advocates include Whewell, Hempel and Reichenbach to name but a few.

All parties agree that this schema is an oversimplification for a number of reasons. Auxiliary hypotheses are usually needed to link theory with empirical data; even then the link may not literally be deductive; theories are rarely abandoned on the basis of a single false prediction; and some true predictions seem to have greater confirmatory value than others. So clearly there is more to be said; but even so, the H-D schema is still widely thought to capture a key part of the scientific method. For there are numerous scientific examples which the schema appears to fit (e.g. in classical astronomy) and

many contemporary scientists claim to employ the H-D method in their work. Even Glymour (1980), a staunch critic of hypothetico-deductivism, admits that it is "obviously the correct account of a great deal of the history of science" (p.48).[1]

The most controversial aspect of the H-D method concerns the inference from a correct empirical prediction to the truth of the theory from which the prediction was derived, which we call "H-D inference". From the viewpoint of deductive logic, H-D inference is of course fallacious; as Salmon (1967) stressed, it is the fallacy of affirming the consequent. So if it is true that scientists often employ H-D inference, or something like it, the question immediately arises whether this is rationally justified. Is it ever rational to take true predictions to confirm a theory, and if so then when, and why? This is a classic question in the philosophy of science, that continues to be discussed today.

Our concern here is with an argument frequently made by philosophers of science of the Bayesian persuasion. They argue that Bayesianism yields an immediate justification of H-D inference. The putative justification derives from a simple point about probability. If theory T implies testable consequence e, i.e. $T \Rightarrow e$, then for any probability function P such that $0 < P(T), P(E) < 1$, it follows that $P(T \mid e) > P(T)$. Therefore, if a scientist's initial (subjective) probability for T lies strictly between 0 and 1, and similarly for e, then upon learning that e is true and updating by conditionalization, they will necessarily become more confident in T. Thus it appears that the core H-D idea – taking correct empirical predictions to confirm a theory – admits of an immediate Bayesian rationale, for it follows from the core Bayesian principle that one should update one's subjective probabilities by conditionalization.

This argument appears repeatedly in the Bayesian literature, often being adduced as a point in favour of the Bayesian approach to confirmation over its rivals. Thus Earman (1992) says that an important "success story" of the Bayesian approach is that it can "explain why... hypothetico-deductive evidence is confirmatory" (p.233). Howson and Urbach (2006) say that a

[1]The H-D schema has been heavily criticized, on the grounds that a deductive link between theory and empirical data is neither necessary nor sufficient for the data to confirm a theory. However, that the H-D schema does not yield a complete analysis of confirmation is compatible with its being a central part of the scientific method. We share the view of a number of recent authors, including Gemes (2005), Sprenger (2011) and Betz (2013), that the H-D schema is right in important respects, despite the traditional objections.

"characteristic pattern of scientific inference occurs when a logical consequence of a theory is shown to be true and the theory then regarded as confirmed". They continue: "Bayes' theorem shows why and under what circumstances a theory is confirmed by its consequence" (p.93-4). In a similar vein, Salmon (2001) says that Bayesianism identifies "the grain of truth in the H-D method" (p.70); while Talbott (2008) says that "one of the most important sources of support for Bayesian confirmation theory is that it can explain the role of hypothetico-deductive explanation in confirmation". Finally, Strevens (2017) says that the Bayesian approach "reproduces the central principle of hypothetico-deductivism and – what H-D never does – gives an argument for the principle" (p.31). The list could easily be extended.

Note that all of these authors take for granted that H-D inference is actually used in science – for otherwise, they would hardly regard it as a point in favour of Bayesianism that it supplies a rationale for such inference. In this article, we similarly take for granted that H-D inference, or something like it, is a central part of the scientific method; our aim is to critically examine the supposed Bayesian justification of it. In short, we argue that the Bayesian justification depends on characterizing H-D inference in a particularly simple way, and this characterization is not true to the actual pattern of H-D inference that we find in science.

In a brief but insightful discussion, van Fraassen (1985) registers two objections to the Bayesians' attempt to justify H-D inference, an attempt which he describes as "useless" (p.284). The first relates to the familiar Duhemian point about the need for auxiliaries in prediction.[2] If, as Duhem taught, the logical relation between theory and evidence is not $T \Rightarrow e$ but rather $T \wedge A_1 \wedge \cdots \wedge A_n \Rightarrow e$, there is no longer any guarantee that learning e will raise the probability of T. There is an extensive Bayesian literature on how to deal with this point.[3] Many Bayesians try to accommodate the Duhemian point by supposing that the scientist has a joint prior over theory and auxiliaries, which, they argue, yields an explanation for why the scientist

[2]Van Fraassen does not explicitly reference Duhem, but rather observes that the evaluation of whether a hypothesis is confirmed when one of its consequences is verified "cannot be without reference to background information" (p.243). Strictly speaking, the Duhemian point is slightly different, namely that a hypothesis will only *have* testable consequences in the presence of suitable background information; however this difference does not matter here.

[3]See in particular Dorling (1979), Howson and Urbach (2006) pp. 92-102, Earman (1992) pp. 83-5, and Strevens (2001).

apportions praise / blame between theory and auxiliaries in the way that they do. Whether this is a real explanation is moot; but in any case, the complications raised by the Duhemian point are clearly well-known to those Bayesians who claim to justify H-D inference.

By contrast, van Fraassen's second objection is much less well-known. He himself makes the objection only in passing, and it receives no attention at all in the recent Bayesian literature. The objection turns on a simple point: what we learn from observation or experiment will often be *logically stronger than what our theory implies*. This point is independent of the Duhemian point above, so for simplicity we may ignore auxiliaries and suppose that the logical relation is given by $T \Rightarrow e$; equivalently, we may suppose that any necessary auxiliaries are part of background knowledge. The point is that in testing the truth of e, what we learn will often be some stronger proposition e' such that $e' \Rightarrow e$ but $e \not\Rightarrow e'$. In other words, we do indeed learn that an empirical consequence of our theory is true, but that is only *part* of what we learn.

This mismatch in logical strength between a theory's testable implication and what is learned empirically may arise for a number of reasons. For example, a theory may predict that two variables will be correlated, but observation reveals a particular degree of correlation. Or a theory may predict that a physical magnitude will lie in a certain interval, while experiment shows that it lies in a narrower interval. Or a theory may predict that a certain phenomenon will occur, while observation reveals that it occurs in one specific way. These are not mere abstract possibilities, as we shall see.

To illustrate, consider the theory (or hypothesis) that humans share a more recent common ancestor with chimpanzees than with gorillas. This, together with standard evolutionary assumptions, implies that humans should exhibit greater genetic similarities with chimps than with gorillas. Now this prediction was borne out by DNA sequencing of the three species' genomes. However the sequence data revealed a very specific pattern of similarity – for example, it showed that what differences there are between humans and chimps are concentrated in a few genomic regions ("human accelerated regions"), which was not predicted (Pollard *et al.* 2006). Thus what was learned empirically, from the sequence data, is that a consequence of the theory is true, but that is only part of what was learned.

This simple logical point complicates the Bayesians' claim to justify H-D inference, as van Fraassen rightly saw, and as we explain in detail below. But its precise significance for that claim needs to be carefully analyzed. That is

the goal of this paper.

We proceed as follows. In Section 2 we spell out in detail the complication raised, for the Bayesian justification of H-D inference, by the mismatch in logical strength between prediction and observation. We illustrate the complication with a toy example then a real example. In section 3, we consider a possible response on the part of the Bayesian, namely that the mismatch in logical strength may be ignored as a "harmless idealization". We show that although this response fails in the general case, it succeeds under certain specifiable conditions; however these conditions have limited scope. In section 4, we turn to examples of successful H-D inference from science. We begin with the famous case of Semelweiss, originally used by Hempel to illustrate the H-D method, and then discuss three cases from modern particle physics. We show how the logical mismatch problem arises in these cases. In section 5, we conclude that the pattern of H-D inference of which there is a ready Bayesian justification is only rarely the pattern actually at work in science.

2 The complication

Suppose it is true that in a typical successful test of a theory, what a scientist learns from observation / experiment is logically stronger than what their theory implies. This has an immediate bearing on the putative Bayesian justification of H-D inference, presuming that the version of Bayesianism in question is the standard one, in which updating by conditionalization plays an essential role.[4] To see why, it is crucial to state the conditionalization rule carefully (as many authors do not). The rule tells an agent how to change their prior probabilities upon learning new information, where "learning new information" means that some proposition about which the agent was previously unsure suddenly gets probability 1. But crucially, this proposition must be the logically strongest thing that the agent learns. So the conditionalization rule does *not* say that if P_t is your probability function at time t and

[4]There is a purely "synchronic" form of Bayesianism that deals with an agent's beliefs at a single point in time, and so has no need for any updating rule. We share Earman's view that "Bayesianism without a rule of conditionalization is hamstrung", as it is too weak to say anything useful about scientific inference (1992, p.161). Moreover, those philosophers who claim that there is a Bayesian justification of H-D inference do not appear to have the conditionalization-free version of Bayesianism in mind.

you learn e between times t and $t+1$, then your new probability function $P_{t+1}(-)$ should equal your old conditional probability function $P_t(- \mid e)$. Rather, it says that this is how P_t and P_{t+1} should relate if e is the logically strongest proposition that you learn between t and $t+1$.

Importantly, the requirement to conditionalize on the logically strongest proposition learned is not an optional extra, or a piece of methodological advice about which there might be a philosophical dispute.[5] Rather it is an integral part of the definition of Bayesian updating. There is good reason for this, since if an agent attempts to conditionalize on only part of what they have learned this quickly leads to contradiction. Thus suppose that the strongest proposition that an agent learns between times t and $t+1$ is X, but instead they try to conditionalize on Y, where $X \Rightarrow Y$ and $Y \not\Rightarrow X$. To reveal the contradiction we need only ask: what is the value of $P_{t+1}(X)$? On the one hand it should equal 1, since the agent has learnt X. But on the other hand, $P_{t+1}(X)$ should equal $P_t(X \mid Y)$, since the agent has conditionalized on Y, which yields contradiction except in the special case where $P_t(X \mid Y) = 1$.

With this point clearly in mind, consider again a circumstance where theory T implies testable consequence e but observation reveals that a stronger proposition e' is true. Again, we assume that the scientist's prior probabilities for T, e and e' lie strictly between zero and one. Now although $P(T \mid e) > P(T)$, we can conclude nothing about whether $P(T \mid e')$ is greater than, less than, or equal to $P(T)$; all of these are possible. This is easily seen from Figure 1; we can vary the value of $P(T \mid e')$, while keeping that of $P(T)$ fixed, simply by moving the position of the set T while keeping its size fixed, so that the hashed area $(T \wedge e')$ constitutes a larger or smaller fraction of the set e'. Therefore when the scientist conditionalizes on what they learn from observation, there is no guarantee that they will become more confident in T. Indeed, it is even possible that $P(T \mid e') = 0$, i.e. that e' conclusively refutes the theory. (In terms of Figure 1, this would mean that T and e' were disjoint.) In short, learning that a theory's empirical consequence is true will not automatically lead a Bayesian agent to become more confident in the theory, unless the truth of the empirical consequence is the logically strongest thing that they learn.

A variant on an example of Rosenkrantz (1982) illustrates how additional

[5]For this reason the requirement should not be thought of as a version of the "principle of total evidence"; for that principle, as usually construed, is precisely a piece of methodological advice.

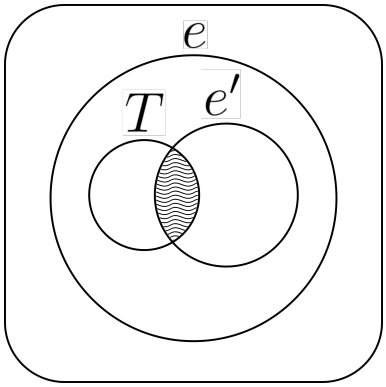

Figure 1: *Logical relations between T, e and e'*

information, over and above what a theory implies, can radically change confirmatory relationships. Five guests are at a dinner party, and each places their hat on a peg by the door. Consider the "theory" T that each guest will leave the party with someone else's hat. A scientist wishes to test this theory, so stands outside the door and asks each departing guest to check whether they have inadvertently taken someone else's hat. Suppose that the scientist learns that the first four guests have indeed left with someone else's hat (e). Since e is a logical consequence of T, learning e confirms T. However suppose instead that the observer learns the stronger proposition e', which says that guest 1 leaves with guest 2's hat, guest 2 with guest 1's, guest 3 with guest 4's, and guest 4 with guest 3's. Clearly, $e' \Rightarrow e$. But notice that e' not merely fails to confirm T but conclusively refutes it (since e' implies that guest 5 leaves with their own hat). Thus in this case, we have $P(T \mid e) > P(T)$ but $P(T \mid e') < P(T)$, since $P(T \mid e') = 0$.

This is a toy example, but it is not difficult to find real examples with a similar structure. Consider the famous Michelson-Morley experiment of 1887. Following Brown (2005), this example can be mapped onto our framework as follows. Let T be the 1880s ether theory given by the Maxwell-Lorentz equations for electromagnetism defined in some inertial frame of reference S, identified as the ether's rest frame. Let e be a core prediction of T, namely that *relative to* S, the two-way light speed in vacuo is constant, in the sense that it is isotropic, independent of the speed of the source and has the value $c \approx 3 \times 10^8$ ms^{-1}. (Brown calls e the "light principle" and quotes Pauli who called it the "true essence of the old aether point of view".) Thus we have that $T \Rightarrow e$. What about e', the experimental finding? According to Brown,

the null result of the Michelson-Morley experiment, that is, the absence of interference fringes, has the "obvious implication" that the two-way light speed in vacuo is constant (in the above sense) in *both* the ether rest frame *and* an approximately inertial frame co-moving with the earth; so we may identify e' with this conjunctive statement. We then have $e' \Rightarrow e$ and $e \not\Rightarrow e'$. Now of course, e' was regarded by many scientists in the period following the experiment, including Einstein, as strongly *disconfirming* theory T, or even refuting it outright. That is, from the Michelson-Morley experiment, scientists learned that a prediction of T was true, but that was only part of what they learned; and the additional information radically changed the evidential situation.

What about the reverse case, in which the logical mismatch between a theory's testable implication (e) and what is learned empirically (e') is in the other direction, i.e. e is logically stronger? This occurs, for example, when a theory predicts that a given physical magnitude has a precise point value but experiment shows that it lies in an interval around that value. Such cases, which are also commonplace in science, are not similarly problematic for the putative Bayesian justification of H-D inference. For it is still true that what is learned empirically is a logical consequence of the theory, since $T \Rightarrow e \Rightarrow e'$. Therefore when the scientist learns e' and conditionalizes, their confidence in T will necessarily increase, though of course by less than if they had learned e. Thus the Bayesian justification works fine in this case.

What is the moral? For all its popularity, the idea that there is a Bayesian justification of H-D inference is less clearcut than it seems. Even in the best-case scenario in which auxiliaries are treated as known, this idea relies on an implicit assumption. It assumes that the pattern of scientific inference in question, whose justification is at issue, can be characterized as "becoming more confident in a theory when one learns that one of its empirical consequences is true and nothing stronger", rather than "becoming more confident in a theory when one learns, *inter alia*, that one of its empirical consequences is true". Only on the former characterization is there an automatic Bayesian justification of the inference pattern. The key questions then, are whether this is an accurate characterization of the H-D inferences actually found in science, and if not, whether it is a legitimate idealization. We address these questions in the next two sections, in reverse order.

Firstly a preliminary worry must be addressed. We have emphasized that a Bayesian agent must always conditionalize on the total information, or logically strongest proposition, that they learn. However, when a scientist

makes an observation or does an experiment, there may be no unambiguous way of saying what exactly the total information *is* that they thereby learn. Consider, for example, the 2015 experiments at the Laser Interferometer Gravitational-Wave Observatory (LIGO), the results of which were published under the title '*Observation* of gravitational waves from a binary black hole merger' (Abbott et al. 2016, our emphasis). What exactly was it that was "observed" in the experiment? The binary black hole merger itself? Transverse waves of spatial strain traveling at the speed of light? That the interferometer's arms had undergone a momentary and minute change in length? That a "chirp" signal of a characteristic frequency, amplitude and duration had been received? The difficulties here are familiar: scientists' observation reports are couched in theoretical language; observation and inference are often partly intertwined; and "raw" observational data needs to be processed and interpreted before it can be related to theory.

This suggests that the notion of "the strongest proposition learned empirically" is not completely determinate – there may be multiple candidates for what this proposition is, and no good way of choosing between them. So perhaps we should conclude that there is simply no objective fact about whether, in a given empirical test, what a scientist learns is or is not stronger than what their theory implies? However, this is surely an overreaction. For after all, there clearly is such a thing as throwing away information. If a scientist observes that the thermometer reading is 20, then the proposition that the thermometer reading is below 25 is only part of what they have learned. We suggest that in many cases, on any "reasonable" account of what exactly was learned from observation, there will be a straightforward answer to the question of whether its content exceeds the theory's implication or not.

3 A harmless idealization?

One possible response to the foregoing argument is this. "The traditional H-D schema – 'T implies e, we learn that e is true, thereby confirming T' – is not intended as a literal description of scientific practice, but rather as an idealization that captures its methodological essence. So although strictly speaking, a scientist may learn more from observation than what their theory implies, it is usually harmless to ignore this."

This response may appear too quick, given that additional information can easily alter confirmatory relations, as we have seen. But on the other

hand, there is something intuitively right about the response, since a maximally specific account of what a scientist learns from observation would include lots of irrelevant detail. Consider the following toy example. A theory implies the empirical prediction: "in test conditions C, the needle on the dial will point to 4". A scientist sets up test conditions C and observes that the needle does indeed point to 4. As they are observing the dial, the fire alarm in the lab goes off. Strictly, the information learned from the act of observation is "the needle points to 4 and the fire alarm has gone off", which is stronger than what the theory implies. But the additional information – that the fire alarm has gone off – is completely irrelevant, we may assume. So for the purposes of methodological reconstruction it seems harmless, indeed preferable, to treat the scientist as having learned the truth of the theory's prediction rather than the stronger proposition that implies it.

To assess this response, we need to probe the "harmless idealization" defence further. Suppose as before that $T \Rightarrow e$, but a scientist learns the stronger proposition e', where $e' \Rightarrow e$ and $e \not\Rightarrow e'$. As before, P is the scientist's initial probability function, and we assume that $0 < P(T), P(e) < 1$. Suppose that, as a deliberate idealization, we treat the scientist as having learned e, the prediction of the theory, rather than e'. Clearly, this idealization will be harmless whenever $P(T \mid e') = P(T \mid e)$. More generally, whenever the quantities $[P(T \mid e') - P(T)]$ and $[P(T \mid e) - P(T)]$ have the same sign, the idealization will be harmless if our interest is only in the qualitative question of whether the scientist's evidence confirms their theory, rather than in the quantitative question of by how much. Now as we know, in general neither of these "harmlessness" conditions need obtain. But we can identify three special cases in which one or both of them will.

3.1 Case 1: Conjunctive evidence

The first case is where the stronger proposition e' is the conjunction of e with something else, i.e. $e' \equiv e \wedge X$. Of course it is always possible to express e' in this form by taking X to be the material conditional $(\neg e \vee e')$. But in some cases, it will be possible to express e' as $e \wedge X$ where X is distinct from $(\neg e \vee e')$. If so, we will say that e' is "conjunctive". More precisely, e' is conjunctive with respect to the probability function P iff the domain of P contains a proposition X, distinct from both e' and $(\neg e \vee e')$, such that $e' \equiv e \wedge X$. Note that when e' is not conjunctive, this may either be because no such X exists, or because the domain of P is insufficiently rich to contain

it.[6]

Intuitively, if e' is conjunctive and if X is an entirely irrelevant piece of information, as in our fire alarm example above, then treating the scientist as having learned e, the prediction of theory T, rather than e', should be harmless. To assess this, consider the following two conditions:

$$P(e \wedge X) = P(e) \cdot P(X) \tag{1}$$
$$P(T \wedge X) = P(T) \cdot P(X) \tag{2}$$

Condition (1) says that X is probabilistically independent of e, while (2) says that X is probabilistically independent of T. These are natural formalizations of the idea that the additional conjunct X contains information that is irrelevant to e, the theory's prediction, and also irrelevant to T, the theory itself.

It is straightforward to show that conditions (1) and (2) jointly suffice for the numerical equality of $P(T \mid e)$ and $P(T \mid e')$.[7] So where the evidence is conjunctive, and these two independence conditions hold, the idealization is entirely harmless, whether we are interested in the qualitative question of whether the scientist's evidence confirms their theory, or in the quantitative confirmation of by how much. This explains what is going on in the fire alarm example above. And it justifies the intuitive thought that, as methodologists, using a maximally specific statement of what a scientist learns from observation or experiment is not always necessary.

To see why the distinction between conjunctive and non-conjunctive evidence matters here, note that if e' is non-conjunctive – that is, can only be expressed in the form $e \wedge X$ by taking X to be the material conditional – then condition (1) is logically unsatisfiable; for the material conditional $(\neg e \vee e')$ cannot be probabilistically independent of e (except in the trivial case where $P(e) = P(e')$). Therefore, while conditions (1) and (2) themselves make no

[6]No such X need exist, since there is no well-defined operation of "subtracting" part of the content of a proposition from it to leave a unique propositional remainder, as Yablo (2015) emphasizes. Note that it is *not* correct to identify the material conditional $(\neg e \vee e')$ as the additional information provided by e' over and above e; the factorization of e' into the conjunction of e and $(\neg e \vee e')$ is one among many such, and not uniquely privileged, as Redhead (1985) observes.

[7]By definition, $P(T \mid e') = P(T \mid e \wedge X) = P(T \wedge e \wedge X)/P(e \wedge X) = P(T \wedge X)/P(e \wedge X) = P(T)/P(e)$ (by conditions (1) and (2)). But $P(T \mid e) = P(T)/P(e)$, since $T \Rightarrow e$. Therefore $P(T \mid e') = P(T \mid e)$.

reference to the conjunctive / non-conjunctive distinction, their joint satisfaction is only possible when e' is conjunctive (again, excluding the trivial case where $P(e) = P(e')$).[8]

What about a case where condition (2) holds but (1) does not (which can occur whether or not e' is conjunctive)? In such a case, $P(T \mid e')$ will not in general equal $P(T \mid e)$. However, given that $T \Rightarrow e$, from condition (2) alone it follows that both $P(T \mid e') - P(T) > 0$ and $P(T \mid e) - P(T) \geq 0$, with equality only in the trivial case where $P(e \wedge X) = P(X)$.[9] So excluding this trivial case, both e and e' confirm T; and thus the idealization will be harmless if we care only about the qualitative question of whether the scientist's evidence confirms their theory. But if we are also interested in degree of confirmation, it will not be harmless. Conversely, if condition (1) holds but (2) does not, the idealization will not in general be harmless, whether we are interested in qualitative or quantitative confirmation.

3.2 Case 2: Interval-valued prediction

Our second case in which it is harmless to treat the scientist as having learned e rather than e' is quite different. It concerns theories that make "interval-valued predictions". Suppose that theory T predicts that some real-valued quantity λ lies in a given interval $[a, b] \in \mathbb{R}$; while experiment reveals that λ lies in a particular sub-interval $[c, d]$ of $[a, b]$ (Figure 2). Such cases arise not uncommonly in scientific practice, and are easily seen to instantiate the pattern described above: what the scientist learns empirically is logically stronger than what their theory implies. Thus we may define $e \equiv \lambda \in [a, b]$ and $e' \equiv \lambda \in [c, d]$, where $T \Rightarrow e$ and $e' \Rightarrow e$.

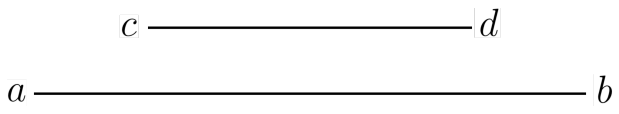

Figure 2: *Interval-valued prediction*

[8]Note that this does not mean that in a non-conjunctive case, the idealization in question cannot be harmless, for conditions (1) and (2) are *sufficient* for $P(T \mid e) = P(T \mid e')$, not necessary.

[9]By definition, and using condition (2) and the fact that $T \Rightarrow e$, we have $P(T \mid e') = P(T \mid e \wedge X) = P(T \wedge e \wedge X)/P(e \wedge X) = P(T \wedge X)/P(e \wedge X) = P(T) \cdot P(X)/P(e \wedge X) \geq P(T)$, with equality only if $P(X) = P(e \wedge X)$.

As before, we have $P(T \mid e) > P(T)$, but nothing follows about whether $P(T \mid e') > P(T)$, nor therefore about whether the scientist will become more confident in T when they conditionalize on e'. However, given two fairly natural assumptions about the prior distribution P, it follows that $P(T \mid e) = P(T \mid e')$, and thus that the idealization is harmless vis-à-vis both qualitative and quantitative confirmation. The assumptions are:

(U_1) conditional on $\lambda \in [a, b]$, λ is uniformly distributed on $[a, b]$
(U_2) conditional on T, λ is uniformly distributed on $[a, b]$

Assumption U_1 means that the prior probability that λ lies in any given sub-interval of $[a, b]$ is proportional to the length of the sub-interval; while U_2 means that this proportionality also holds conditional on the truth of T. Plausibly, P will satisfy these two assumptions if the scientist has no prior information about where in the interval λ lies; and if theory T provides no information about this either. For then, the scientist may well consider that λ is no more likely to lie in one sub-interval of $[a, b]$ than in another of equal length; and that this will continue to be so on the assumption that theory T is true. We show in Appendix 1 that these two uniformity assumptions jointly imply that $P(T \mid e) = P(T \mid e')$.

Clearly, assumptions U_1 and U_2 represent applications of the classical principle of indifference, and as such are no better or worse than other applications of that principle. A proponent of the indifference principle would hold that if the scientist is entirely ignorant about where in the interval $[a, b]$ λ lies, and if theory T says nothing about this, then they are rationally required to satisfy U_1 and U_2. Most contemporary Bayesians would be reluctant to insist on this, given the well-known problems that arise from unrestricted application of the classical indifference principle.[10] But whether rationally required or not, the point here is simply that *if* the scientist's prior P satisfies these two uniformity assumptions, for whatever reason, then it makes no difference to their posterior credence $P(T \mid e')$ which sub-interval of $[a, b]$ experiment reveals λ to lie in. That is, even though what the scientist learns empirically is stronger than what the theory implies, it is harmless to ignore this and to treat the scientist as having learned the weaker statement that $\lambda \in [a, b]$.

Importantly, conditions U_1 and U_2 are sufficient for $P(T \mid e) = P(T \mid e')$,

[10]The best known of these is Bertrand's paradox; see Gyenis and Rédei (2014) for a recent treatment.

not necessary. One might wonder whether, in the case of interval-valued prediction, we can find necessary and sufficient conditions for this equality to hold? In a trivial sense, we clearly can. Suppose that the following condition holds:

$$\frac{P(\lambda \in [c,d])}{P(\lambda \in [a,b])} = \frac{P(\lambda \in [c,d] \mid T)}{P(\lambda \in [a,b] \mid T)} \tag{3}$$

That is, the ratio of the prior probability that λ lies in $[c,d]$ to the probability that it lies in $[a,b]$ remains unchanged when we condition on T. It is easy to verify that this is necessary and sufficient for $P(T \mid e) = P(T \mid e')$ (recalling the definitions $e \equiv \lambda \in [a,b]$ and $e' \equiv \lambda \in [c,d]$). More generally, if and only if the same holds true not just for $[c,d]$ but for *every* sub-interval of $[a,b]$, then the idealization will be harmless irrespective of which sub-interval λ is found to lie in experimentally.

Though necessary and sufficient, condition (3) is not especially useful, since it is really no more than a simple re-arrangement of the equality $P(T \mid e) = P(T \mid e')$. By contrast, our uniformity conditions U_1 and U_2, are more useful, for we can give a plausible if not completely compelling argument, based on the indifference principle, for why a scientist's prior P might satisfy them; but the price for this is that the conditions are merely sufficient for the idealization to be harmless, not necessary.

Finally, one might wonder whether, in the case of interval-valued prediction, we can isolate a sufficient condition for $[P(T \mid e') - P(T)]$ and $[P(T \mid e) - P(T)]$ to have the same sign, and thus for the idealization to be harmless vis-à-vis qualitative confirmation alone? Unlike in the case of conjunctive evidence, no non-trivial condition suggests itself.

3.3 Case 3: Disjunctive prediction

Our third case is in effect a discrete analogue of case 2, but merits separate discussion as the scientific contexts in which it applies are different. Suppose that theory T makes a prediction of disjunctive form, that is, of the form $[e_1 \vee e_2 \vee \cdots \vee e_n]$, where the e_i are pairwise exclusive. Through observation, a scientist learns that one particular disjunct, say e_2, is true; again, this is stronger than what the theory implies. So in terms of our previous notation, $e \equiv [e_1 \vee e_2 \vee \cdots \vee e_n]$ and $e' \equiv e_2$. A natural interpretation is this: the theory predicts that a given phenomenon will occur but leaves open certain aspect(s)

of how it will occur; observation reveals that the phenomenon occurs in one particular way.[11]

As before, we know that $P(T \mid e) > P(T)$, but it is open whether or not $P(T \mid e') > P(T)$. However, suppose that the scientist's prior P satisfies two conditions. Firstly, they regard each of the e_i as equally likely; secondly, they also regard the e_i as equally likely conditional on the truth of T. That is:

(U_3) $\quad P(e_i \mid e) = 1/n$ for all i
(U_4) $\quad P(e_i \mid T) = 1/n$ for all i

Again, conditions U_3 and U_4 represent standard applications of the indifference principle. If the scientist has no information about which disjunct will occur, and if theory T says nothing about the matter, then plausibly, they will regard all n disjuncts as equi-probable, and also as equi-probable conditional on T. It is straightforward to show that U_3 and U_4 suffice for the equality of $P(T \mid e)$ and $P(T \mid e')$, and thus for the idealization to be harmless: the scientist's posterior credence, on learning the truth of the one of the disjuncts, will be exactly the same as if they had learned merely that the disjunction is true.[12]

As in the interval-valued case, it is controversial whether a scientist who knows nothing about which disjunct will occur is rationally *compelled* to satisfy U_3 (and similarly for U_4). But in practice, relying on the indifference principle in discrete cases is fairly common, in both everyday and scientific reasoning. As before, we do not need to take a stand on this issue here. Our point is simply that, if a scientist's prior does satisfy U_3 and U_4, for whatever reason, then the idealization in question is harmless.

Finally, as in the interval-valued case, note that conditions U_3 and U_4 are sufficient for the equality of $P(T \mid e)$ and $P(T \mid e')$, not necessary. Again, a necessary and sufficient, though trivial, condition for this equality is simply that $P(e_i \mid e) = P(e_i \mid T)$ for each i. That is, the proportion of the total probability of e that P gives to disjunct e_i should remain unchanged by

[11] Note that we could trivially treat every case where $e' \Rightarrow e$ as disjunctive, by writing $e \equiv [e' \vee (e \wedge \neg e')]$. We could exclude this, if desired, by requiring that each of the disjuncts must be strictly stronger than $(e \wedge \neg e')$. However there is little point to this exclusion, since conditions U_3 and U_4 are satisfiable even if we take e to be $[e' \vee (e \wedge \neg e')]$. In this respect, the logic of the situation is quite different from that of conjunctive evidence (case 1), above.

[12] An analogue of the argument in Appendix 1 shows this.

conditioning on T. The uniformity conditions U_3 and U_4 simply represent one natural and obvious way of ensuring that this is so.

3.4 Summary

This section has explored the suggestion that applying the simple H-D schema ("T implies e, we learn that e is true, thereby confirming T") to a situation where in reality, a scientist learns the truth not of e but of a stronger proposition e', can be justified as a harmless idealization. In general it cannot, since there is no guarantee that $P(T \mid e)$ will equal $P(T \mid e')$, nor even that $[P(T \mid e') - P(T)]$ and $[P(T \mid e) - P(T)]$ will have the same sign. However we identified three special cases which suffice for the idealization to be harmless. The first is where e' is conjunctive and the independence conditions (1) and (2) are satisfied. The second is where e is an interval-valued prediction and the uniformity conditions U_1 and U_2 are satisfied. The third is where e is a disjunctive prediction and the uniformity conditions U_3 and U_4 satisfied.

These results go some way towards defending the putative Bayesian justification of H-D inference from the objection outlined in Section 2; they show that for the purposes of methodological analysis, it is not always necessary to use a maximally specific description of what a scientist learns empirically. But the results are fairly limited in scope, and seem unlikely to cover all the examples in which a scientific theory is regarded as confirmed when one of its empirical consequences is verified. We turn next to an examination of some examples.

4 Scientific examples

In Section 2, we saw that the putative Bayesian justification relies upon characterizing H-D inference as "becoming more confident in a theory upon learning that one of its empirical consequences is true and nothing stronger". In this section, we examine whether this characterization is true to scientific practice. We argue that very often, it is not.

Consider the case that Hempel originally used to illustrate H-D inference in science: Semelweiss's study of the causes of puerperal fever in the Vienna General Hospital in the 1840s (Hempel 1966). As Hempel reconstructs the story, Semelweiss considered various hypotheses for why the rate of puerperal fever was higher in one maternity division, staffed by doctors,

than in another, staffed by midwives. Eventually he hit on the hypothesis that cadaveric matter on the doctor's hands was the cause of puerperal fever (midwives did not perform autopsies). To test this, he required that doctors wash their hands in chlorinated lime immediately after performing an autopsy, predicting that this would lead the rate of puerperal fever in the doctor-staffed division to fall. The prediction was borne out, thus confirming Semelweiss's hypothesis.

As Hempel describes it, the Semelweiss story instantiates the simple H-D schema: "$T \Rightarrow e$, we learn that e is true, thus confirming T". But this is clearly a considerable oversimplification.[13] Though the rate of puerperal fever did indeed fall after the hand-washing measures were introduced, the rate did not fall to zero but continued to fluctuate, remaining higher in the hospital than outside it; and sporadic outbursts still occurred (Tulodziecki 2013). So if e is the proposition that the rate of fever will fall after hand-washing is introduced, then although Semelweiss did indeed learn the truth of e, this was only part of what he learnt; and the additional information was not a logical consequence of his hypothesis.

The moral of the Semelweiss example likely generalizes to other cases where a causal model is subjected to empirical test. For example, a simple non-parametric causal model of the form $X \rightarrow Y$ implies that intervening on variable X will alter the value of Y; but what we learn when we do the experiment is a specific statement about the values of Y for different values of X. So again, what is learned empirically is stronger than what the causal model implies. The same is true when more complex causal models are tested, whether by experimental intervention or by comparing the model's predictions against observational data.

These examples are suggestive but not decisive. For if a hypothesis under test yields a fairly broad-brush prediction, to the effect that changing one variable will have some effect on another variable, it is almost inevitable that, in a successful test, what is learned empirically will be logically stronger than that the prediction is true. It is hard to see how Semelweiss could have learned *only* that hand-washing reduces the incidence of puerperal fever, without learning anything about the pattern and magnitude of the reduction. So a more telling example would involve a theory which makes a highly

[13]To be fair to Hempel, he does acknowledge that his description of the Semelweiss case is oversimplified, though for reasons other than the one emphasized here. Tulodziecki (2013) argues that the Semelweiss case is rather more complex than its standard portrayal in the philosophy literature.

specific prediction; as then, it will not be a foregone conclusion that what is learned empirically exceeds what the theory implies. The obvious place to seek such examples is physics. In particular, the field of particle physics is characterized by a combination of quantitatively precise novel prediction and high-precision experimentation, which makes it the ideal testing ground for the Bayesian justification of H-D inference.

4.1 Higgs scalar boson

Our first example is particle physics' most famous recent prediction: the Higgs boson. Following Dawid (2015), the crucial elements of the case are as follows. In the 1960s it was proposed by Higgs and others that a particular mechanism for "spontaneous symmetry breaking" based upon a massive scalar boson (i.e. the Higgs particle) would provide an explanation for particle masses, and would thus be the final piece in the jigsaw of the "standard model" of particle physics.[14] In and of itself, the theory of the Higgs mechanism does not put any constraint on the mass of the relevant particle: any non-zero mass Higgs particle could play the relevant role in the spontaneous symmetry breaking mechanism. However, when combined with data from previous experiments, certain broad features of the Higgs model imply an upper bound on the Higgs mass of 141 GeV at the 95% confidence level (where GeV is giga electron-volts, and the speed of light is set to one).[15] Previous experiments also imposed a lower bound on the Higgs of 115 GeV, at the 95% confidence level. The Higgs was detected at the LHC in 2012 with a mass between 125 and 127 GeV.

Dawid (2015) offers an in-depth discussion of how the prior constraints on the Higgs mass affected the evidential connection between the LHC experiment and the Higgs model. Whilst the details of his analysis are orthogonal

[14]The particular problem is that unbroken gauge symmetry does not allow for the observed massive vector bosons, like $W^{\pm}$ and Z^0, that mediate the weak nuclear force, nor for the mass spectra of fermionic matter, like quarks and electrons, that are the constituents of atoms. The Englert-Brout-Higgs-Guralnik-Hagen-Kibble spontaneous symmetry breaking proposal is that, due to the existence of a massive scalar boson, although the theory's Lagrangian remains gauge symmetric, the electroweak gauge symmetry is broken at the theory's ground state, thus allowing for massive vector bosons and fermionic particles. See Dawid (2015, p.77) and Friederich (2014) for more details.

[15]The relevant feature here is the ability of "virtual" Higgs particles to contribute to scattering cross sections in experiments at energies below the level needed to produce genuine Higgs particles. See Dawid (2015, pp. 84-85) for details.

to our concerns, his conclusion is highly relevant. If, as Dawid suggests, we adopt what he calls the "theoretician's perspective" on the Higgs discovery, and take the prediction being tested to be not just the existence of the Higgs but also the bounds 115–141 GeV, then the case neatly illustrates our point. Let T be the theory of the Higgs mechanism, and assume that previous experimental results (prior to 2012) were part of background knowledge. Let e be the prediction that the Higgs exists and has a mass between 115–141 GeV. Let e' be the result of the LHC experiment, namely that the Higgs mass lies between 125–127 GeV. So we have $T \Rightarrow e$, $e' \Rightarrow e$ but $e \not\Rightarrow e'$, i.e. the experimental finding is logically stronger than what the theory implies.

Two points merit emphasis here. Firstly, the Higgs example is surely a paradigm of H-D inference in science: a theory makes a prediction which is put to empirical test, the prediction is correct, and the theory is thereby confirmed. Secondly, the 2012 experiment was widely regarded as a scientific triumph, and had a tangible impact on physicists' confidence in the correctness of the theory. The fact that the experimental finding was stronger than what the theory implied in no way dented the physicists' conviction that the standard model had received striking empirical confirmation. Now as we have seen, many philosophers argue that H-D inference admits of a simple Bayesian justification. Since the Higgs case involves H-D inference, then if these philosophers are right, we should expect the Bayesian justification to apply here. But it does not: since e' is stronger than e, learning e' need not increase scientists' confidence in T.

Might the "harmless idealisation" defence be invoked at this point? The Higgs example is a case of interval-valued prediction, and as we saw in §3.2, in such cases the two uniformity conditions U_1 and U_2 jointly suffice for the equality of $P(T \mid e)$ and $P(T \mid e')$. Do these conditions apply in the Higgs case? Arguably not. U_1 says that the prior probability that the Higgs mass lies in any sub-interval of 115–141 GeV is proportional to the size of that sub-interval (given relevant past experimental knowledge). U_2 says that this proportionality also holds conditional on the theory of the Higgs mechanism. Now although the issue is complex and not fully settled, there is a general consensus that different possible values of the Higgs mass between 115–141 GeV place different constraints on the possible physics "beyond the standard model". In particular, before the 2012 experiments, it was widely taken to be a fairly generic prediction of low energy supersymmetric extensions to the

standard model that the Higgs mass should not be much above 140 GeV.[16] Thus physicists' background beliefs plausibly meant that they had a non-uniform prior distribution over the interval 115–141 GeV. There is no reason why conditioning on T should induce uniformity, so it seems that both U_1 and U_2 fail in this case.

Of course, U_1 and U_2 are sufficient rather than necessary conditions for the harmless idealization defence to apply, so the defence might go through via some other route. But the onus is on the Bayesian to establish this. On the face of it, the Higgs case is a paradigm of successful H-D inference in science, but one to which the putative Bayesian justification of that inference pattern does not apply.

Importantly, we are not arguing that no possible Bayesian reconstruction of the scientists' reasoning in the Higgs case could be given. Quite possibly it could, as Dawid (2015) has suggested. However, such a reconstruction would necessarily have to go beyond imputing to scientists the simple H-D schema ("T implies e, we learn that e is true, therefore confirming T.") For this schema is untrue to the Higgs case, given the logical mismatch between prediction and observation; and thus the Bayesian justification of that schema does not apply here.[17]

4.2 $W^{\pm}$ vector boson

Our next example is the discovery of the $W^{\pm}$ vector boson. This is a type of force-carrying particle, the mass of which the Higgs scalar boson was called in to explain. Together with the Z^0 vector boson, the $W^{\pm}$ is a key part of the standard model explanation for nuclear decay processes such as Beta-decay. In particular, the $W^{\pm}$ plays a key role in mediating fermonic interactions within the Glashow-Weinberg-Salam theory of unified electroweak interactions. The force-mediating role of the $W^{\pm}$ in electroweak theory is directly analogous to the role of the photon in mediating electromagnetic interactions in quantum electrodynamics. Unlike the photon, however, the $W^{\pm}$ is a massive gauge boson. In fact, the value of its mass plays a crucial theoretical role in that it is directly constrained by electroweak theory.

Following Franklin (1986, pp. 170-2), we can reconstruct the key steps in the prediction and discovery of the $W^{\pm}$ as follows. By the early 1980s,

[16] See for example Espinosa et al. (1993) and Zhang et al. (2008).

[17] Thanks to Richard Dawid and Radin Dardashti for discussions on this issue.

electroweak theory was well corroborated even though the $W^{\pm}$ had not been detected.[18] Since the $W^{\pm}$ was a core part of electroweak interaction, there was enough interest in detection to spur the construction of a new "Super Proton Synchrotron" (SPS) Collider at CERN with the primary goal of detecting the particle. The report on the experiment was published in 1983 and indicated a missing transverse energy of the correct magnitude for $W^{\pm}$ decay. From the SPS experiments, the "observed" value of the $W^{\pm}$ mass was calculated and found to be in excellent agreement with the value predicted by electroweak theory. Furthermore, Franklin notes, both the number of observed production events and the transverse momentum distribution of the $W^{\pm}$ particles were in good agreement with theory. Again, this is a paradigm of the H-D method: a theoretical prediction is put to empirical test, the prediction is correct, and the theory thereby confirmed.[19]

To apply our framework, let us take T to be electroweak theory, e to be the predicted value of the $W^{\pm}$ mass, and e' the experimentally determined mass. Now the theoretical prediction, in this case, is not a point-value but rather the interval 82 ± 2.4 GeV. It might seem odd that the theory only fixes one of its fundamental parameters to an interval. But in this case, as in many others in physics, the extraction of a quantitative prediction from the theory involves both mathematical approximation techniques and reliance upon prior experimental data, which is why electroweak theory does not yield a point-valued prediction for the $W^{\pm}$ mass. What about the experimentally-determined mass? As in the Higgs case, we again find that the experiment yields an interval-valued result.[20] In the $W^{\pm}$ case, the experimental determination of the mass was 81 ± 5 GeV, an interval that entirely encompasses the predicted interval. We thus have that $e \Rightarrow e'$ but $e' \not\Rightarrow e$. So the logical mismatch is exactly the opposite of the Higgs case: what was learnt empirically was weaker, not stronger, than what electroweak theory implies.

In §2, we noted that when e' is logically weaker than e, the standard

[18] In fact, Glashow, Weinberg and Salam had already been jointly awarded the 1979 Nobel prize for developing electroweak theory based upon the discover of the Z^0 at CERN in 1973.

[19] Franklin notes that in these experiments, the agreement with theory was also taken by scientists to validate the observations. This illustrates the Duhemian point made in §1: the theory-observation relation is often more complex than as depicted by the simple H-D schema. Again, however, we leave this complication aside here.

[20] This is usually the case, since even well-validated experimental procedures can only resolve a value to a finite degree.

Bayesian justification of H-D inference works fine: since $T \Rightarrow e \Rightarrow e'$, learning e' will necessarily increase the probability of T, though by less than if e had been learnt. Since the experiment in question did indeed lead physicists to become more confident in electroweak theory, the $W^{\pm}$ case may seem to support those who argue that H-D inference in science admits of a simple Bayesian justification.

Taken on its own, the $W^{\pm}$ case does indeed support the Bayesians' argument. However, when taken in conjunction with the Higgs case, matters look different. In both cases, a theoretical prediction was put to experimental test, the prediction was verified, and scientists' confidence in the respective theory increased. But the logic of the two cases is fundamentally unalike, for in the Higgs case e is logically stronger than e', while in the $W^{\pm}$ case the reverse is true. And as we know, this makes a big difference from a Bayesian point of view. So if the Bayesians' story is correct, we should expect scientists to see a fundamental distinction between these two cases in respect of their status as instances of qualitative confirmation. But this is not what we find. On the contrary, in *both* cases the scientific consensus was that the experimental findings had confirmed the theory.

That scientists' judgments of whether the evidence qualitatively confirms the theory are insensitive to the relative strength of e and e' casts doubt on the putative Bayesian justification. Moreover, in the $W^{\pm}$ case it is clearly a *contingent fact* that the experimentally-determined mass interval was larger than the predicted interval, for the former depends on the degree of experimental precision possible at the time. Imagine that a global recession had delayed the construction of the accelerator needed to detect the $W^{\pm}$ by twenty years. Technological advances would almost certainly have meant that the precision of the experiment was greatly increased, so that the experimentally determined interval could have lain inside the predicted interval. If so, the logical relation between e and e' would be inverted, and the simple Bayesian story would no longer apply. However, it is hard to believe that the scientific community would regard the resulting evidential situation as fundamentally different. In this counterfactual scenario, no less than in the actual scenario, a theory makes a correct prediction and is thereby confirmed.

In fact, we do not need to invoke counterfactual scenarios to demonstrate the insensitivity of scientists' confirmatory judgments to the direction of logical mismatch between e and e'. For experimental precision and mathematical approximation techniques can both evolve over time, so it sometimes happens that the observed and predicted intervals are successively and independently

restricted. We turn to such a case next.

4.3 Fine Structure Constant

The fine structure constant α characterizes the strength of the electromagnetic interactions described by quantum electrodynamics, and its measurement is a remarkable example of successful quantitative prediction. In fact, the measurement of α is the basis for the much-heralded claim that quantum electrodynamics is the most precisely tested theory in the history of science.[21] Recently, α has been subject to hyper-precision measurements with accuracy of the order of 10^{-10} (Bouchendira *et al.* 2011). Again, we have what looks like the classical H-D inference pattern. The theory yields a prediction, the prediction matches the observation, and the theory is thereby confirmed.

Is the case of α akin to that of $W^{\pm}$, where e is logically stronger than e', or to the Higgs, where the reverse is true? Revealingly, we in fact find *both* of these logical structures at different points in time. Over the last forty years, the precision of the experimental measurement of α has sometimes been greater than the precision of the theoretical calculation; while at other times the reverse has been true. This is because better and better approximation techniques for evaluating the relevant integrals in quantum electrodynamics have evolved in tandem with the refinement of the relevant experimental techniques.[22]

This presents an interesting test case for the Bayesian justification of H-D inference. Let T denote quantum electrodynamics, e_t the predicted interval for α at time t, and e'_t the experimentally determined interval at time t. Both e_t and e'_t are time-indexed, as explained above. Thus it is possible that for some times t, $e'_t \Rightarrow e_t$ and $e'_t \not\Rightarrow e_t$ while for other t, $e_t \Rightarrow e'_t$ and $e_t \not\Rightarrow e'_t$. And this is in fact what has happened. Over the last forty years, the relative logical strength of e_t and e'_t has flipped back and forth again and again, due to a combination of factors including technological development, experimental funding, computational power and scientific ingenuity.

[21] In fact, strictly speaking, the measurement of α is not a test of quantum electrodynamics since it is an unconstrained parameter in the theory. Rather, two alternative values of α are derived and compared via the experimental measurement and theoretical determination of the anomalous magnetic moment of the electron. These complications do not alter the basic inferential structure of the case.

[22] See Figure 1 of Bouchendira *et al.* (2011) for details. The current best predicted value is given in Aoyama *et al.* (2012).

Now, if the putative Bayesian rationale for H-D inference were right, we should expect the confirmational significance of the experimental findings to depend on whether or not theoretical precision was ahead of experimental precision at the time in question. At times when e_t was logically stronger than e'_t, we should expect the experimental findings to definitely increase scientists' confidence in quantum electrodynamics; while at times when e'_t was stronger, we should not automatically expect this. However, in fact this distinction carried no such significance. So far as we know, there is nothing to suggest that scientists took experimental confirmation of quantum electrodynamics to depend on the contingent circumstance of whether e_t or e'_t was logically stronger.

Might the "harmless idealization" defence be invoked at this point? For example, perhaps scientists at times when e'_t was stronger than e_t were implicitly appealing to the uniformity conditions U_1 and U_2 of §3.2? This is certainly possible. But it is equally possible that, as in the Higgs case, U_1 and U_2 did not apply, for example because scientists' background beliefs led them to expect that the value of α was more likely to be at the upper or lower end of the predicted interval.[23] The onus is thus on the Bayesian to explain why, if their analysis of H-D inference is correct, scientists' judgments were insensitive to the relative logical strength of e_t and e'_t, and why, in cases where e'_t was stronger, scientists nonetheless took the experimental findings to support quantum electrodynamics.

5 Conclusion

Many Bayesians claim to be able to justify H-D inference, which they adduce as a major selling point of their approach over its rivals. van Fraassen (1985) dismissed this putative justification as "useless" on two grounds: firstly because it ignores the role of auxiliaries in deriving predictions; and secondly because it ignores the fact that what is learned empirically is often logically stronger than what the theory implies. The first point has been extensively discussed in the literature, but the second completely ignored. Our aim here has been to fill this lacuna.

[23] An example of such a background belief would be scepticism regarding specific parts of the code used for the calculation and thus an expectation that future refinements will lead to a value at the lower or upper end of the interval.

Since a Bayesian agent must always conditionalize on the logically strongest proposition they learn, it follows that the putative Bayesian justification of H-D inference relies on characterizing it in a particular way, namely "becoming more confident in a theory when one learns that one of its empirical consequences is true and nothing stronger". The key questions then, are whether this is an accurate characterization of the H-D inferences actually found in science, and if not, whether it can be viewed as a harmless idealization of the methodologist?

We have found that the answer to both questions is "no". In general, if $T \Rightarrow e$ but a scientist learns the stronger proposition e', then it is not harmless to treat the scientist as having learned e. We showed that under specific conditions, $P(T \mid e)$ and $P(T \mid e')$ will indeed be equal, but there is no reason to think that these conditions usually obtain in practice. Our case studies suggest that it is quite common for experimental data to be logically stronger than the theoretical prediction being tested; indeed this occurs in paradigm instances of H-D inference in science. Moreover, scientists' judgments about evidential support are insufficiently insensitive to the relative logical strength of theoretical prediction and experimental data to be consistent with the rationale for H-D inference that the Bayesians offer.

In conclusion, while "useless" is perhaps too strong, our analysis suggests that the putative Bayesian justification of H-D inference has much less explanatory power than many philosophers think. The pattern of inference of which there is a ready Bayesian justification is not the pattern that one finds actually at work in science. Whatever its other virtues, the idea that Bayesianism offers a simple justification of a pervasive pattern of scientific inference appears to have been oversold.

Acknowledgments

We are deeply appreciative of feedback received on a draft manuscript from Erik Curiel, Radin Dardashti, Richard Dawid, and Elliott Sober. We also received valuable comments from an audience at a 'Work in Progress' talk in Bristol.

Appendix 1

By assumption U_1:

$$P(\lambda \in [c,d] \mid \lambda \in [a,b]) = \frac{P(\lambda \in [c,d])}{P(\lambda \in [a,b])} = \frac{d-c}{b-a}$$

By assumption U_2:

$$\frac{P(\lambda \in [c,d] \mid T)}{P(\lambda \in [a,b] \mid T)} = \frac{d-c}{b-a}$$

Therefore:

$$\frac{P(\lambda \in [c,d])}{P(\lambda \in [a,b])} = \frac{P(\lambda \in [c,d] \mid T)}{P(\lambda \in [a,b] \mid T)}$$

Cross-multiplying:

$$\frac{P(\lambda \in [a,b] \mid T)}{P(\lambda \in [a,b])} = \frac{P(\lambda \in [c,d] \mid T)}{P(\lambda \in [c,d])}$$

Multiplying across by $P(T)$:

$$\frac{P(T) \cdot P(\lambda \in [a,b] \mid T)}{P(\lambda \in [a,b])} = \frac{P(T) \cdot P(\lambda \in [c,d] \mid T)}{P(\lambda \in [c,d])}$$

Applying Bayes' theorem:

$$P(T \mid \lambda \in [a,b]) = P(T \mid \lambda \in [c,d])$$

Applying definitions $e \equiv \lambda \in [a,b]$ and $e' \equiv \lambda \in [c,d]$:

$$P(T \mid e) = P(T \mid e')$$

Q.E.D.

References

Aoyama, T., Hayakawa, M., Kinoshita T., and Nio, M. (2012) Tenth-order QED contribution to the electron g-2 and an improved value of the fine structure constant. *Physical Review Letters* 109(11): 111807.

Abbott et al. (2016) Observation of Gravitational Waves from a Binary Black Hole Merger. *Physical Review Letters* 116(6): 061102

Betz, G. (2013) Revamping Hypothetico-Deductivism: a Dialectic Account of Confirmation. *Erkenntnis* 78, 991–1009.

Bouchendira, R., Cladé, P., Guellati-Khélifa, S., Nez, F. and Biraben, F. (2011) New Determination of the Fine Structure Constant and Test of the Quantum Electrodynamics. *Physical Review Letters* 106(8): 080801.

Brown, H. R. (2005) *Physical Relativity.* Oxford: Oxford University Press.

Dawid, R. (2015) Higgs Discovery and the Look Elsewhere Effect. *Philosophy of Science* 82(1), 76–96.

Draper, P. and Rzehak, H. (2016) A review of Higgs mass calculations in supersymmetric models. *Physics Reports*, 618(11), 1—24

Dorling, J. (1979) Bayesian personalism, the methodology of scientific research programmes, and Duhem's problem. *Studies in History and Philosophy of Science Part A* 10(3): 177–187.

Earman, J. (1992) *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory.* Cambridge MA: MIT Press.

Espinosa, J. R., & Quirs, M. (1993). Upper bounds on the lightest Higgs boson mass in general supersymmetric standard models. *Physics Letters B*, 302, 51-58.

Franklin, A. (1986) *The Neglect of Experiment.* Cambridge: Cambridge University Press.

Friederich, S. (2014) A philosophical look at the Higgs mechanism. *Journal for General Philosophy of Science* 45, 335—350 .

Gemes, K. (2005) Hypotheico-Deductivism: incomplete but not hopeless. *Erkenntnis* 63(1): 139–47.

Glymour, C. (1980) Theory and Evidence. Princeton: Princeton University Press.

Gyenis, Z. and Rédei, M. (2014) Defusing Bertrand's Paradox. *British Journal for the Philosophy of Science* 66(2): 349–73.

Hempel, C. (1966) *Philosophy of Natural Science*. Englewood Cliffs, NJ. Prentice-Hall.

Howson, C. and Urbach, P. (2006) Scientific Reasoning: the Bayesian Approach, 3rd edition. La Salle: Open Court.

Pollard, K. S., Salama S. R., King B., Kern, A. D., Dreszer, T. Katzman, S., *et al.* (2006) Forces Shaping the Fastest Evolving Regions in the Human Genome. *PLoS Genetics* 2(10): e168. https://doi.org/10.1371/journal.pgen.0020168.

Redhead, M. (1985) On the Impossibility of Inductive Probability. *British Journal for the Philosophy of Science* 36(2): 185–91.

Rosenkrantz, R. (1982) Does the Philosophy of Induction Rest on a Mistake? *Journal of Philosophy* 79: 78–97.

Salmon, W. (1967) *Foundations of Scientific Inference*. Pittsburgh: University of Pittsburgh Press.

Salmon, W. (2011) Explanation and Confirmation: a Bayesian Critique of Inference to the Best Explanation. In G. Hon and S.S Rakover (eds) *Explanation: Theoretical Approaches and Applications*, pp. 61–92, Dordrecht: Springer.

Sprenger, J. (2011) Hypothetico-deductive confirmation. *Philosophy Compass* 6(7): 497–508.

Strevens M. (2001) The Bayesian Treatment of Auxiliary Hypotheses. *British Journal for the Philosophy of Science* 52: 515–537.

Strevens, M. (2017) Notes on Bayesian Confirmation Theory. http://www.nyu.edu/classes/strevens/BCT/BCT.pdf (Unpublished notes).

Talbott, W. (2016) Bayesian Epistemology. *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), url =https://plato.stanford.edu/archives/win2016/entries/epistemology-bayesian/.

Tulodziecki, D. (2013) Shattering the Myth of Semmelweis. *Philosophy of Science* 80(5).

van Fraassen, B. (1985) Empiricism in the Philosophy of Science. In P. Churchland and C. Hooker (eds) *Images of Science*, Chicago: University of Chicago Press.

Yablo, S. (2014) *Aboutness*. Princeton: Princeton University Press.

Zhang, Y., An, H., Ji, X., & Mohapatra, R. N. (2008). Light Higgs mass bound in supersymmetric left-right models. *Physical Review D*, 78(1), 011302.