The Other Kind of Confirmation

(aka In Praise of Instance Confirmation)

Michael Strevens

Draft of April 2006

Abstract

It is argued that the relation of instance confirmation has a role to play in scientific methodology that complements, rather than competing with, a modern account of inductive support such as Bayesian confirmation theory. When an instance confirms a hypothesis, it provides inductive support, but it also provides two things that other inductive supporters normally do not: first, a connection to "empirical data" that makes science epistemically special, and second, inductive support not only for the hypothesis as a whole, but for its parts. Further, when it is conceived in the right way, instance confirmation can duck the arguments most often thought to refute it. A causal account of instantiation, thus of instance confirmation, is offered that looks to deliver on all of the foregoing promises.

1. Introduction

Of accounts of scientific confirmation, Bayesianism is surely the current champion, while instantialism is widely regarded as a hopeless relic. I want to make a case for instantialism, not as a rival, but as a complement to Bayesian confirmation theory (or if Bayesianism should fall, as a complement to whatever theory of confirmation comes to play Bayesianism's present role).

I will not merely be pointing out that, on the Bayesian or any other view, a hypothesis's confirmers usually include among their number its positive instances. Everyone knows this already.

I will rather argue that there are two quite different, though systematically related, inductive relations between evidence and theory, either of which might properly be called *confirmation*. Bayesianism provides a rather good account of the weaker of these two relations; a certain sort of instantialism provides, I think, the correct account of the stronger.

The two relations are not in any way incommensurable: on the contrary, they are interlocking parts of a single inductive system, and indeed, if things go well, the stronger relation is to be defined partly in terms of the weaker relation.

I call the relations B-confirmation and I-confirmation. The 'B' is for Bayesianism and the 'I' for instantialism, but these are just mnemonics: nothing about the notion of B-confirmation presupposes the involvement of subjective probability, and nothing about the notion of I-confirmation presupposes the involvement of instantiation. Bayesian confirmation theory should, then, be thought of as a particular account of B-confirmation, and instantialism as a particular approach to understanding I-confirmation.

This paper is structured around two considerations that give us reason—so I argue—to posit the existence of two quite different kinds of confirmation relation. The first consideration arises from Hempel's argument that confirmation flows along the entailment relation from entailer to logical consequence (section 3). The second focuses on a narrow but supremely important sense in which evidence can be *empirical* (section 4). After presenting these considerations, I sketch an instantialist account of the stronger variety of confirmation relation, that is, of I-confirmation. Let me begin with the briefest of overviews of the Bayesian and instantialist approaches to confirmation.

2. Two Theories of Confirmation

On the instantialist theory of confirmation—Hempel (1945) is my paradigm—there are two ways that a hypothesis can be confirmed. First, hypotheses are directly confirmed or disconfirmed by their instances—if they are deterministic, they are of course confirmed by their positive instances and disconfirmed by their negative instances.[1]

Second, hypotheses are confirmed or disconfirmed if they stand in the right sort of logical or semantic relation to a hypothesis that has been directly confirmed or disconfirmed. According to Hempel's consequence rule, for example, when a set of hypotheses is confirmed, their logical consequences are also confirmed. All confirmation is triggered by instance confirmation, then, but a hypothesis can be confirmed by a piece of evidence that is not one of its instances, by way of a *confirmation transmission rule* such as the consequence rule.

Hempel's account is perhaps best known for its satisfiability theory of instantiation, on which a hypothesis of the form *All Fs are G* has as an instance, and thus is confirmed by, an observation of an object that is either not *F* or is both *F* and *G*. This account will play no role in the present paper.

On the Bayesian approach to confirmation, a hypothesis h is confirmed by a piece of evidence e just in case the receipt of e raises the subjective probability of h, that is, just in case

$$\Pr(h|e) > \Pr(h)$$

The evidence e disconfirms h just in case $\Pr(h|e) < \Pr(h)$.

A note on evidence: I follow Hempel and the Bayesians in taking the evidence, in what follows, to be an observation statement, that is, a proposition

1. If they are statistical, their confirmation or disconfirmation is a matter of comparing the ratio of positive to negative instances with the ratio predicted by the hypothesis. The question how to go about this comparison is not one on which I take a stand in what follows.

spelling out some aspect of the observable world—the blackness of a particular raven, the value of particular readout, and so on. The evidence might equally be regarded as the state of affairs represented by the observation statement; this fits well with an instantialist approach, since it is natural to take the blackness of the raven itself, not a statement of such, as what instantiates *All ravens are black*. What evidence is not is some object in all its particularity, such as an individual black raven. That is, though the evidence may be an observed worldly state of affairs—the blackness of a particular raven—it cannot be the raven itself. The same goes, in what follows, for instances. It is a raven's combined ravenhood and blackness (or an observation statement that specifies the coinstantiation of these properties) that instantiates the hypothesis, not the raven itself.

3. Inductive Support versus Inductive Saturation

When certain aspects of Hempel's motivation for adopting an instantialist view are properly understood, I will argue, it is apparent that Hempel is seeking to explicate a kind of inductive relation that is quite distinct from the relation formalized by Bayesian confirmation theory.

Instantialism requires what I called in the previous section rules for the transmission of confirmation. Hempel (1945) considers in particular two possible rules, the consequence rule and the converse consequence rule.[2] Notoriously, the rules cannot be endorsed simultaneously, and Hempel opts for the consequence rule (described in section 2). The details, though germane to the present discussion, are omitted here for reasons of space. My principal interest is Hempel's justification of his decision to opt for the consequence rule:

> Any . . . consequence [of a hypothesis] is but an assertion of all or part of the . . . content of the original [hypothesis] and has therefore

2. The rules are first introduced as conditions of adequacy for any account of confirmation, but Hempel's intention is to put them to use as transmission rules; in his formulation of instantialism, the consequence rule is the one and only rule of transmission.

> to be regarded as confirmed by any evidence which confirms . . . the latter (Hempel 1945, 31).

In this passage Hempel supposes that, when a piece of evidence confirms a hypothesis, it provides evidential support for every part of the hypothesis.

To a Bayesian, this constitutes an oversight of Oedipal proportions: it is not only possible, but extremely common, for a piece of evidence to confirm a hypothesis by inductively supporting one, but not all, of its parts. To see this, simply take evidence e that confirms hypothesis h, then choose another hypothesis g that is intuitively irrelevant to e and h. (You might formalize this irrelevance by requiring that g be independent of h conditional on e; see Fitelson (2002).) On the Bayesian approach, confirmation is probability-raising. Since e raises the probability of h, it also raises the probability of hg, for an irrelevant g. Informally, if your confidence in h increases, and g has nothing to do with h, then your confidence in hg ought to increase, since you are surer than before of the first conjunct and no less (or more) sure of the second. Thinking of hg as a single hypothesis, then, you have a case where e confirms a hypothesis by confirming one part of the hypothesis but not the other.

How could Hempel have failed to recognize this possibility? Has the introduction of probabilistic thinking to confirmation theory so clarified the subject matter that what seemed compelling to the ancients is immediately apparent to us as an elementary error?

This is hardly plausible, considering that it was Hempel himself who introduced the problem of irrelevant conjuncts to confirmation theory. I propose that what Hempel intended his instantialism to capture was an evidential relation to which the presupposition inherent in his justification of the consequence rule applies. In Hempel's idiolect, a piece of evidence cannot be said to confirm a hypothesis unless it provides inductive support for all its parts. This is, obviously, a stronger notion of confirmation than the Bayesian version, on which evidence can confirm a hypothesis by providing support for only one of its parts.

What we have, then, are two classes of inductive relation. I call the first and weaker relation *inductive support*. It is the kind of relation that Bayesians attempt to explicate. In what follows, at any rate, I take inductive support and probability increase to amount to more or less the same thing, so that a piece of evidence inductively supports a hypothesis just in case it increases the probability of the hypothesis. Evidence may inductively support a hypothesis, then, without supporting all of its parts.

The second and stronger relation I call *inductive saturation*. A piece of evidence inductively saturates a hypothesis just in case it inductively supports all of its parts, and thereby supports the hypothesis as a whole.[3]

While the Bayesian notion of confirmation is a support relation, the confirmation notion pursued by Hempel is, I suggest, a saturation notion. Further, instantialist approaches to confirmation ought to be interpreted as, and are generally quite suitable for, explicating a saturation relation. The instantialist confirmation relation sketched in section 5, in particular, ought to be regarded as such. The difference between B-confirmation and I-confirmation is in part, then, that the latter but not the former is a relation of saturation; for the other differences, see sections 4 and 5.

Why think that inductive scientific reasoning needs something more than inductive support? Why think that confirmation theory should be concerned with inductive saturation, so much so that the label *confirmation* might just as well be applied to saturation relations as to relations of support?

Before I give an answer, let me emphasize that to be concerned with saturation is not to be unconcerned with support. Science needs an inductive support relation at least as much as it needs a saturation relation—obviously so, since saturation is characterized in terms of support. In what follows I make a case for the absolute, not the relative, importance of saturation: the newfound status of saturation should in no way be regarded as detracting from the status of support. My goal, recall, is not to argue that the Bayesian theory fails to capture the true nature of confirmation, but rather to argue that there are two kinds of inductive relations, each scientifically important enough in itself that it might be called *confirmation*. Bayesianism gives a good account

3. Support of all the parts is not quite sufficient for support of the whole. The clause after the *thereby* is not merely a clarification, then, but a part of the definition of saturation.

of one of these relations; my point is that we need to attend carefully to the other, as well.

What, then, is the special significance of saturation? I take my aim here to be to show that scientists have a considerable interest in knowing that a piece of evidence not only supports, but saturates, a hypothesis.

Scientific inference aims to do two things with the evidence. First, the results of an empirical test should tell us, as far as possible, what to expect from other such tests. Second, the results of an empirical test should tell us what hypotheses and theories to believe, or to "accept".

Consider all of the hypotheses and theories that predict (either by entailing or by probabilifying) a given empirical datum. Given the two aims of science, it is important to know

1. Which of the hypotheses satisfy the following condition: because of the evidence, we have better reason than before to believe all the other experimental predictions made by the same hypothesis, and

2. Which of the theories satisfy the following condition: because of the evidence, we have better reason than before to believe all the other parts of the same theory, and therefore, in time, to accept the theory as a whole.

What we want to know, then, is that the evidence inductively saturates the hypothesis and the theory (using an intuitive but uncontroversial notion of parthood).

Why is support on its own not enough? To be told that the evidence supports, but not that it saturates, a hypothesis, is to be told that you have more reason to believe the hypothesis than before, while leaving open the question whether you have reason to rely any more than before on some, or perhaps many, of the hypothesis's predictions. Such information is nearly useless to you in pursuit of the first aim of science. Equally, to be told that the evidence confirms, but not that it saturates, a theory, is to be told that you have more reason to believe the theory than before, but to leave open the question whether you have any more reason than before to believe some, or perhaps many, of the theory's parts. Such information is nearly useless to you in pursuit of the second aim of science. (Compare Glymour (1980).) To assess the ultimate scientific significance of a piece of evidence, then, you need to know which hypotheses and theories it saturates.

I am not saying, of course, that information about support is useless on its own. Quite the contrary: after all, enough of the right kind of information about support will entail the facts about saturation. My point is simply that the evidence has a special significance for the theories it saturates, as opposed to the theories it supports without saturation; science, then, has a special interest in saturation.

I am also not saying that we can do with information about saturation alone. Sometimes we must make decisions that turn on hypotheses that are supported by some of the evidence, but not saturated by any of the evidence; in such cases, we of course care about relations of unsaturating support. In any case, it barely even makes sense to claim that we can get by with saturation alone, given that the facts about saturation are in part constituted by the facts about support.

The picture I have sketched so far is very simple. Suppose that you define a relation of inductive saturation as suggested above:

> *e* inductively saturates *h* just in case *e* inductively supports all of *h*'s parts

using some preexisting inductive support relation, perhaps the Bayesian relation. The issue of parthood aside, the structure of the inductive saturation relation is inherent in the structure of the support relation. The two are different relations, but the saturation relation as it were reduces to the support relation (more exactly, to the support relation plus the facts about parthood).

You might even say that, although confirmation theory aims at two goals, a support relation and a saturation relation, its method is, just as the Bayesians implicitly suppose, the elucidation of single relation, namely, the relation of inductive support. That would be to go too far. Even given all the facts about support, there are interesting new things to say about how and when saturation emerges from support, that is, about which sorts of evidence tend to saturate which sorts of hypotheses, and why. Compare higher and lower level sciences: even if all the biological facts are entailed by all the physical facts, there are still many interesting things to say about the way in which the entailment works. Philosophers may disagree on whether such an entailment relation exists, but they agree that even if it does, there is still work for the biology department to do.

However, I do not want to make a big deal of this. Although it is a mistake to overstate the simplicity of the picture I have sketched, it is nevertheless unquestionably quite straightforward: B-confirmation is inductive support, and I-confirmation is simply saturating B-confirmation.

The situation becomes more complicated in the next section, where I propose that I-confirmation—the relation that Hempel and others seek to capture using the notion of instance confirmation—is characterized in terms of a narrower kind of inductive support than B-confirmation.

4. Scientific Exceptionalism

Science is different—and better, if only at figuring out what is really going on. Why? It might be white coats and expensive machinery. It might be the sheer intellectual power of its practitioners. It might be a culture of "organized skepticism". It might be an undying commitment to the truth. It might be a mix of these things.

But at least as important, in the estimation of many philosophers, is that science is especially sensitive, and in the right sort of way, to the "empirical data" (using *empirical* in a sense narrower than is usual in philosophy). Other systems of knowledge production—the interpretation of sacred texts, the construction of conspiracy theories, perhaps certain aspects of folk medicine, and of course philosophy—may make arguments, both inductive and deductive, as sophisticated as those found in science; they may be, in some cases, just as well funded; their practitioners may have equally subtle minds; yet they are by comparison with science abject failures because they do not assign the central role that science does to empirical testing.

You might hope that the theory of confirmation would provide an account of this "empirical testing" that makes science so exceptional in both senses of the word. The Bayesian notion of confirmation will, in that case, disappoint you. Bayesianism aims to elucidate the rules of inductive inference wherever it is found. Thus the Bayesian's inductive support relation is instantiated wherever there is competent inductive (and indeed, deductive) argumentation in any of the fields of human inquiry enumerated above. How could it not be? Whatever you think of Biblical scholarship, the unmasking of the Bavarian Illuminati, or counterpart theory, you must admit that they make use of inductive argument. The practitioners of these pursuits can be assigned coherent subjective probability functions, and their reasoning can be interpreted and evaluated as conditionalization according to the probabilities assigned. The same is neither more nor less true of science. Bayesian confirmation theory is thus not in a position, and does not aspire to be in a position, to distinguish on logical grounds between the kind of reasoning found in science and the kinds found in other pursuits.

Let me elaborate, focusing now on science alone. A scientist may come to believe (or accept, or assign a high subjective probability to) a hypothesis for many different kinds of reasons. She may (as noted above) find the hypothesis asserted by a reliable textbook, or by a reliable colleague. She may infer the hypothesis, deductively or inductively, from some other hypothesis or hypotheses she believes. Or she may put the hypothesis to the empirical test. All of these procedures are legitimate sources of knowledge, but it is the last of them that makes science special. Empirical evidence, whether produced by experiment or gathered by fieldwork, is the epistemic lifeblood of science, and the other sources of knowledge just mentioned are respected only insofar as they lie on the path from empirical data to conclusion, and so have a claim to carry the lifeblood themselves.

A natural way to capture this common belief about the nature and the success of the scientific method is to posit a special kind of inductive support, which I will call *empirical inductive support*. A paragraph in a textbook and an experimental test may both lend inductive support to a hypothesis, but only the former can give the hypothesis *empirical* support, or equivalently, only the former support relation belongs to that subclass of the inductive support relations that are the empirical support relations. (I will remind you again that the sense of *empirical* employed here is narrower than the usual philosophical sense, on which any a posteriori inference, including inference from the printed word, is empirical.)

I make the following proposals and claims:

1. Empirical support is important enough that we should consider coining a sense of the term *confirmation* on which evidence *e* confirms *h* only if *e* gives *h* empirical inductive support. In accord with my preexisting terminological liberality, this use of *confirmation* may coexist with other senses, such as the Bayesian's B-confirmation.

2. The instantialists should be interpreted as aiming to give a theory of empirical support. Their key doctrine: a deterministic hypothesis is empirically supported by its positive instances (and empirically counter-supported, as it were, by its negative instances).

3. The notion of empirical support may be profitably combined with the notion of inductive saturation to provide a notion of empirical inductive saturation, defined in the obvious way: *e* empirically inductively saturates *h* just in case *e* provides empirical support for all parts of *h*. The relation of I-confirmation that I am pursuing in this paper is a kind of empirical inductive saturation.

These proposals, you will see, generate the complication promised at the end of section 3: they entail that I-confirmation is a certain, restricted kind of inductive saturation, a kind defined in terms of an inductive support relation distinct from B-confirmation—namely, empirical support.

Even if you accept that the Bayesian's confirmation relation is not a relation of empirical support alone, you might wonder whether Bayesianism will one day be replaced by some other theory of inductive support that offers a more empirical confirmation relation. I suggest not. It is not due to a quirk or weakness of Bayesian confirmation theory that its notion of confirmation embraces non-empirical support. It is due, rather, to Bayesianism's largely successful realization of its ambition to provide what might be called a universalistic conception of inductive support, that is, a conception that is instantiated wherever correct inductive argumentation is found. Because some correct inductive reasoning is not empirical, no evidential relation can be both narrowly empirical and universalistic.

What is the nature of empirical support, and why is it so important to science? I will not provide an entirely general answer to this question, for the following reason: the particular kind of instantialism I advocate in this paper is based on a subclass of empirical support that I call causal-empirical support, and I wish to restrict my attention to this subclass. In the next section, after introducing my instantialism, I will have a characterization of causal-empirical support, and I will say a few words about its epistemological significance.

Two remarks before I turn to instantialism proper. First, you might try to capture what is especially "empirical" about science without positing a special relation of empirical support. A Bayesian, for example, might point to the important Bayesian convergence theorems to demonstrate that enough of the right sort of evidence will create a convergence of opinion. Such a move cannot, I think, succeed. The train of evidence that brings about the convergence might just as well be, for all the mathematics of convergence cares, the utterances of concordant prophets as a stream of empirical data. (I would say more, but space is at a premium.)

Second, note that Hempel's stated motivations for instantialism include a concern with empirical support as well as a concern with inductive saturation—or so it can be argued. Because the argument depends in part on properties of empirical support discussed in the next section, I defer the topic until then.

I have identified two properties of evidential relations that are of particular interest to the scientific enterprise, and that are distinct from the Bayesian confirmation relation: the property of being a relation of inductive saturation, and the property of being a relation of empirical support. The instantialist notion of confirmation, I propose, attempts to unite these two, by offering an evidential relation that combines saturation and empirical support in the obvious way: evidence *e* instance confirms a hypothesis *h* only if *e* empirically supports all parts of *h*. (Note that the converse does not hold: instance confirmation does not capture every kind of empirical support, or at least, I will not try to argue that it does.)

Taking a step back, I have now provided characterizations of two distinct notions of confirmation. On the one hand, you have a confirmation relation that amounts to a universal inductive support relation, that is, the relation that obtains wherever a good inductive argument is found. This is what I call B-confirmation, in honor of Bayesianism's claim to provide an especially powerful theory of universal inductive support. On the other hand, you have a confirmation relation that is a kind of saturating empirical support, which I call I-confirmation both for historical reasons—it is the kind of relation that I suppose Hempel was aiming to capture with his instantialist theory of confirmation—and for more forward-looking reasons, namely, my claim that a certain account of instance confirmation, distinct from Hempel's, does a very good job of capturing I-confirmation.

Let me sketch a simplified version of the instantialism that I favor.

5. A New Instantialism

5.1 Instantiation

What follows is a sketch of an attractive but unsophisticated account of instantiation. The account is restricted in scope. It applies only to the instantiation of causal hypotheses, and in its current formulation, only to causal hypotheses of the form *All Fs are G*. This is not, I think, so great of a narrowing of focus as it seems. Most scientific hypotheses are causal. For example, the raven hypothesis—*All ravens are black*—is not the simple empirical generalization that Hempel took it to be. Properly understood, it does not say merely that everything that is a raven is also black. Rather, it says that there is a certain mechanism, present in or in the vicinity of all ravens, or at least all normal ravens, that causes blackness. (I do not require that the property of ravenhood itself do the causing; that would be to ask too much (Strevens manuscript c).)

I assume that we associate with every causal hypothesis a single mechanism in virtue of which the something about *F* causes *G*-ness. Our description of that mechanism may be radically incomplete, but we do have in mind some sort of characterization of the mechanism, however rudimentary. Call this description the hypothesis's associated causal model.

The effect of the associated model is to turn a hypothesis into a small, self-contained causal theory. Observe that, even if a hypothesis is correct in the overt causal claim it makes, it may be defective in the sense that its underlying theory—its associated causal model—fails to reflect the causal facts: either it misrepresents the true mechanism, or there is no single true mechanism at all, but rather several mechanisms all having the same effect.

Now the account of instantiation. For simplicity's sake, I will talk as though it is things that instantiate hypotheses, though as noted in section 2, it is more accurate to say that states of affairs or propositions are the instances.

An object is a positive instance of a hypothesis of the form *All Fs are G* just in case

1. It is *F* and *G*, and

2. The associated causal model's conditions for operation are satisfied.

By the model's conditions for operation, I mean a certain specified set of conditions necessary for the modeled mechanism to cause *G*-ness. These are what are sometimes called ceteris paribus conditions; they include background conditions for the mechanism's operation, conditions that rule out any interference with the operation, and conditions that rule out the later reversal of the mechanism's output, that is, that specify that no later process destroys the *G*-ness of the *F* in question. (It is somewhat awkward to call this last kind of requirement a condition for the mechanism's operation; rather, it is a condition required for the sustenance of the end-product of the mechanism's successful operation; an example is provided below.)

The set of operation conditions specified by a model, thus required for instantiation, need not be in any sense complete. Why not? Among other things because, were the conditions of operation to be necessary and sufficient for the operation of the mechanism, there could be no negative instances. A negative instance of the canonical hypothesis *All Fs are G* is an *F* for which the conditions of operation obtain, but which is not *G*. Clearly, there could be no negative instances if the operation conditions' obtaining entailed the presence of *G*. It is a tricky question what the relation is between the conditions for operation and the causal model itself. It is on this very issue, indeed, that the simple account of instantiation presented above departs from what I believe to be the true account (Strevens manuscript a).

The account of instantiation I have offered differs from Hempel's satisfiability account (section 2) in two respects. First, and more shallowly, it allows only *F*s to be instances, where Hempel counts all non-*F*s as instances as well. Second, it includes the causal condition, condition (2) (which presupposes the *F*-ness of the instance).

Let me say something more about this condition (2). A black raven, you will see, instantiates the raven hypothesis (on a causal interpretation) just in case its blackness is possibly due to the causal model associated with the law. A raven dyed black is therefore not an instance of the law. Why? While there is a mechanism that is responsible for the blackness of a dyed raven, it is very different from the mechanism described by the raven hypothesis's associated causal model: even the most naive ornithological theorist has in mind a mechanism in which the causation is an internal, biological matter.[4] Technically, what has gone wrong in this case is that a certain condition required for the operation of the natural raven blackness mechanism has been violated, a condition of the sort that requires that the mechanism's operation not be later undone or reversed.

Looking ahead to confirmation, that the dyed raven is no instance is quite encouraging, since you would not count the color of such a bird as evidence for the law. Likewise, a raven bleached white is not a negative instance of the law, and so is not counted as evidence against the law.

A white shoe, I am sure you can see, does not even begin to satisfy the criteria for instantiating the raven hypothesis; its color clearly has nothing to do with the hypothesis's associated causal model. Why no paradox? Hempel's route to the raven paradox assumes the logical equivalence of the hypotheses

4. There is good psychological evidence for this claim; see Strevens (2000).

All ravens are black and *All non-black things are non-ravens*. On the causal gloss given here, these hypotheses mean quite different things: the former makes a claim about the causation of blackness in ravens, while the latter makes a (patently false) claim about the causation of non-ravenhood in non-black things. (Although a white shoe is not an instance of the raven hypothesis, it will in certain circumstances provide some slight inductive support for the hypothesis, by the usual Bayesian arguments.)

I define direct instance confirmation in the same way as Hempel (section 2): a deterministic hypothesis is directly confirmed by its positive instances and directly disconfirmed by its negative instances. You will recall that Hempel then uses a transmission rule, the consequence rule, to allow the confirmation brought by a piece of evidence to percolate from directly confirmed hypotheses to other, logically related, propositions (specifically, the logical consequences of the directly confirmed hypotheses). I will discuss transmission shortly. I first want to vindicate my claim that if e directly confirms h—if e is an instance of h in the sense just described—then e empirically supports all parts of h.

I begin by showing that an instance *saturates* the hypothesis, that is, that it inductively supports all parts of the hypothesis, given the appropriate notion of parthood. I then ask, in a separate section, whether the support offered to the parts by an instance is *empirical*.

5.2 Saturation

An instance of a hypothesis saturates the hypothesis if it provides inductive support for all parts of the hypothesis. What are the parts of a hypothesis? In arguing for the importance of saturation in section 3, I suggested that scientists are particularly interested in two classes of hypothesis "parts": first, a hypothesis's various predictions, and second, the various sub-hypotheses of which it is composed.

In the case of causal hypotheses, I interpret these parts (with something of a stipulative flourish) as follows. The predictions of a causal hypothesis are, loosely speaking, its instances. More exactly, the prediction of a hypothesis such as *All Fs are G* is a conditional claim: any *F* for which the associated causal model's conditions of operation are satisfied will be *G*. The instances are, as it were, not the predictions themselves but the realizations of the predictions—but I will not strive to maintain this distinction.[5] The sub-hypotheses of a causal hypothesis are the propositions that together make up the description of its putative underlying mechanism, that is, its associated causal model. (If it strikes you as peculiar to count the model as a part of the hypothesis, I think you are right. As foreshadowed above, it would be better to call the model a theory, and to think in terms of the predictions and sub-hypotheses of the theory. But having written in terms of hypotheses throughout the paper so far, I prefer to avoid a sudden terminological switch.)

An instance of a hypothesis saturates the hypothesis, then, if first, it provides inductive support for the hypothesis's being instantiated in any other situation to which it applies, and second, it provides inductive support for all parts of its casual model. Let me show that, intuitively, saturation does occur—intuitively, because I am not working with any particular theory of inductive support.

Begin with the parts of the associated causal model. I make the following important assumption: everything mentioned in the model plays an essential part in the putative causal processes that result in the hypothesis's instantiation. For example, every part of the model for the blackness mechanism in ravens needs to be present, according to the story told by the model, for ravens to get their color, or at least raises the probability of their doing so, in an appropriately causal way. As I have put it elsewhere, the parts of the model must all be *difference-makers* (Strevens 2004, manuscript b).

Now, given that the causal model describes only difference-makers for the instance—given that, had any of the the causal factors mentioned by the model not been present, the hypothesis would not have been instantiated, or at least, would have had a lower probability of being instantiated—the fact

5. To elucidate my concern: *All ravens are black* does not predict that any particular object will be both a raven and black, but rather predicts of whatever particular objects are ravens, that they will be black. It is for this reason strictly speaking incorrect to say that any particular instance is per se predicted by the hypothesis.

of instantiation gives us more reason than before to think that the factors are present, just as the model asserts. Thus an instance provides inductive support for all parts of the model, as desired.[6] (An important objection to this line of reasoning will be considered shortly.)

Next, the saturation of the "parts" of a hypothesis that are its predictions. An instance of the causal hypothesis *All Fs are G* gives you reason to believe, by the argument in the previous paragraph, in the existence in or around *F*s of a certain mechanism, namely, the mechanism described by the hypothesis's associated causal model. As such, it lends empirical inductive support to any of the phenomena that such a mechanism tends to generate, or in other words, to the hypothesis's instances.

Let me consider an apparent difficulty with the argument for saturation: the same piece of evidence that gives positive inductive support to an instance or model-part for the reasons given above may also give it negative support for some other reason, sufficient to cancel out the positive support. For example, a piece of evidence might be an instance of two rival hypotheses that agree in some of their predictions but not in others. By providing inductive support to opposing pairs of predictions, the evidence will provide net inductive support to neither. Or (perhaps less likely) the same factor whose presence is stipulated by the causal model for one hypothesis may have its absence stipulated by the causal model for another, so that evidence that is an instance of both hypotheses is on balance neutral as to the factor's presence.[7]

There are two ways to deal with this problem. First, you might relax the criteria for saturation somewhat (at least in the context of the definition of I-confirmation), saying that evidence that I-confirms a hypothesis must empir-

6. Compare and contrast Glymour (1980)'s technique for distinguishing parts of a theory that are confirmed by a piece of evidence; whereas I hold that the parts of a theory describing the causal difference-makers are confirmed, Glymour offers a purely syntactic criterion for the same.

7. There are, for example, two well known mechanisms causing thrombosis in women. One involves the birth control pill; the other involves pregnancy, and thus the absence of the pill. The fact of thrombosis in a particular patient therefore provides positive support for the patient's taking the pill along one path, but negative support along the other. The overall level of inductive support depends, according to the usual Bayesian formula, on both the strength of the support along the two paths and the base rates—essentially, the ratio of pregnant women to women on the pill in the population from which the patient is drawn.

ically support most of its parts, or must empirically support its parts all other things being equal.

Second (taking off, in effect, from the *ceteris paribus* hedge just suggested), you could pay closer attention to the different senses in which one proposition might be said to inductively support another. The standard Bayesian notion of inductive support is an "all things considered" notion, on which evidence *e* supports *h* only if the sum total of its inductive relevance to *h* is positive. There may be more than one "inductive path" from *e* to *h*—*e* may be relevant to *h* in more than one way, as in the examples above. An all-things-considered measure of support aggregates the support over all the paths. (The aggregate may not be a simple sum; see section 6.)

To say that *e* inductively supports *h* is not necessarily to say that *e* supports *h* all things considered. It can also mean that *e* supports *h* along at least one inductive path, without ruling out the possibility that the support is undone by negative support along other paths. Thus it is not inconsistent to say: *e* gives you reason to believe *h*, but it also gives you reason to disbelieve *h*; as a result, *e*'s net impact is neutral or unfavorable to *h*. I propose that the inductive support that an instance invariably lends a hypotheses' parts is of this path-specific variety. Then, I claim, the argument for saturation goes through.

In order to clarify matters, I will incorporate the notions of all-things-considered and path-specific support into the definitions of B-confirmation and I-confirmation, saying that *e* B-confirms *h* just in case *e* provides all-things-considered inductive support for *h*, whereas *e* I-confirms *h* just in case *e* provides path-specific empirical support of some sort for all parts of *h*. However, there exist broader notions of B-confirmation and I-confirmation that do not impose these restrictions;[8] my narrowing is primarily for expository purposes.

To sum up, then: if *e* is an instance of *h*, then *e* inductively saturates *h*, or more exactly, *e* provides inductive support, in the path-specific sense, for

8. For example, there may be a useful notion of B-confirmation on which that fact that *e* B-confirms *h* implies only that *e* provides path-specific support for *h*. What makes it B-confirmation is that first, it does not imply saturation, and second, it is universal—wherever there is inductive support there is B-confirmation. Since all-things-considered support implies path-specific support, this notion of B-confirmation is strictly broader than the official notion defined in the main text.

all parts of *h*, meaning its predictions and the components of its associated causal model. Now to discuss the sense in which the support is empirical.

5.3 Empirical Support

What is empirical support, and why is it important? I will not offer a definition of empirical support. I take it that the instances of a hypothesis are its paradigmatic empirical supporters: if there is such a thing as empirical support at all, then black ravens empirically support *All ravens are black*, black holes empirically support *Black holes exist*, and so on. Although eventually, someone will have to define empirical support and show that instances are empirical supporters, this is very far from being an urgent problem.

Let me therefore simply stipulate that the instances of a hypothesis empirically support the hypothesis, and that they do so because of their intimate causal relation to the hypothesis. Call this *causal-empirical* support. There may be other varieties of empirical support; I put the matter aside for now.

The stipulation allows me to move immediately to what I take to be a more interesting question: why is empirical support of special interest to, and apparently especially desirable to, science? What, in particular, is good about what I am calling causal-empirical support?

The question might be divided into two parts. First, why is causal-empirical support a kind of inductive support? That is, why does a evidence that causal-empirically supports a proposition also inductively support (in a path-specific way, of course) that proposition? Second, what is special about this particular variety of inductive support? What does causal-empirical support have that other varieties of inductive support—at least the non-empirical varieties—do not have?

I will not say much about the first question. There are as many ways to show that causal-empirical support is inductive support as there are theories of inductive support. A Bayesian, for example, can show that an instance of a hypothesis will, as well as raising the probability of the entire hypothesis, raise the probability of the parts of the causal model (ceteris paribus), and so raise the probability of the predictions of the causal model (again ceteris paribus). In a Bayesian or any other demonstration of the inductive power of instances, the key fact is that the parts of the causal model jointly entail, or at least physically probabilify, the instances.[9]

I should add that this sort of Bayesian derivation is not completely straightforward, because it involves splitting the change in the probability of the hypothesis part given the instance into different components, only one of which reflects the path-specific support given to the hypothesis part by the instance in virtue of its being an instance of that particular hypothesis. The ceteris paribus hedges therefore pass over some fancy derivational footwork. For an example of the technique of splitting an all-things-considered probability change into path-specific components, see Strevens (2001).

To the second question, then: why is causal-empirical inductive support somehow preferable to other, non-empirical forms of inductive support? The answer, I think, is that causal-empirical support does not depend on auxiliary hypotheses. This gives the causal-empirical support relation a considerable, and valuable, degree of freedom from the local epistemic context. A particular degree of causal-empirical support is no stronger, inductively, than the same degree of non-empirical inductive support—how could it be?—but has a certain kind of objectivity that the scientific enterprise rightly values.

How is causal-empirical support context free? Begin with a kind of support that is clearly not context free. Thinking David Lewis to be a great philosopher, I am inclined to believe anything I read in *On the Plurality of Worlds*. Thinking David Lewis to be a brilliant but perverse philosopher, you give endorsement in the book no weight at all (perhaps even negative weight). We agree which hypotheses are asserted in *On the Plurality of Worlds*, but we do not agree on the inductive significance of the appearance for the hypotheses themselves, because that significance depends entirely on an auxiliary hypothesis con-

9. It is not quite this simple, in the statistical case. For a Bayesian, to support a hypothesis h or its parts, an instance must be probabilified by h to a greater degree than it is by the competing hypotheses, *weighted by prior subjective probability*. A likelihood approach to statistical inference finds a way to ignore these weights (by understanding support in a comparative way). More generally, the nature of your favorite theory of inductive support will determine the caveats and riders you will want to attach to the claim that instances provide path-specific inductive support for their statistical hypotheses' parts.

cerning the truth-tracking powers of Lewis's philosophical faculties. To put it another way, all paths from the evidence (appearance in the book) to the hypotheses themselves goes by way of a further, quite distinct hypothesis.

Instances are not like this. There is an inductive route from an instance to its hypothesis, and to the hypothesis' parts, that does not make any detours, essentially because a hypothesis' associated causal model is sufficient in itself, without auxiliaries, to entail or to physically probabilify its instantiation. (Reminder: on what it means to probabilify an instantiation, see note 5.) Thus though we may disagree on a lot, we can agree on the causal-empirical support that an instance gives a hypothesis in a way that we cannot agree on the support given a hypothesis by its assertion in *On the Plurality of Worlds*. (Caveats shortly.)

It is this agreement, and its objective basis in entailment/physical probabilification, that gives causal-empirical support its value. Science is a social enterprise. It cannot proceed efficiently without a high degree of consensus on certain things, most of all consensus on the evidence and on the inductive significance of the evidence.[10] Now, if something like the Bayesian theory is correct, there cannot be perfect agreement on the all-things-considered significance of the evidence. It is possible, however, for scientists to selectively ignore, in public though perhaps not in private, the stickily subjective strands of the inductive web in favor of the strands whose strength is equally appreciated by all. What I am calling causal-empirical support is a category of just such strands—the most important category, I would say. If I am right, then the facts about causal-empirical support constitute the principal intersubjective—indeed, the principal objective—basis for the ongoing scientific evaluation of theories.

As I remarked earlier, Hempel's instantialism seems to have been motivated in part by a similar concern. In arguing against hypothetico-deductivism (he calls it the prediction criterion for confirmation), he complains that many of a theory's predictions, which all, according to hypothetico-deductivism, are

10. On some other matters, disagreement may be a virtue: see Kitcher (1993), chap. 8 on the importance of cognitive diversity in science.

confirmers, go by way of auxiliary hypotheses. Hempel implicitly rejects a common defensive reply, that what is confirmed is a conglomerate of theory and auxiliaries. He wants the theory itself confirmed by the evidence, but not by way of auxiliaries. Thus he insists that only instances can directly confirm. (Hempel's dislike of auxiliaries is not obviously motivated by the disadvantages of auxiliaries I have cited here; you would not expect to find him exclaiming that science is a social enterprise. But I conjecture that it is, in the end, based on similar concerns.)

I mentioned some caveats on the objectivity of the empirical support relation. Even when confirmation is based around empirical support, it cannot be entirely context free. One reason, which applies to Bayesians (but not, say, to likelihood theorists), is explained in note 9. Another is that freedom from auxiliaries applies only to the inductive relation between instance and hypothesis. But often auxiliaries are required to infer that you have observed an instance at all. Why? Consider a causal hypothesis of the form *All Fs are G*. To infer that you have an instance, you must infer that you have an *F* that is *G*, and for which the associated mechanism's conditions of operation held. If *F*, *G*, or any of the conditions spelled out in the associated model are not directly observable, you will likely need auxiliary hypotheses to infer their obtaining. Even if the inference from instance to hypothesis is context free, then, the inference from observation to instance may not be.

Hempel avoids this problem by giving a theory that applies to hypotheses couched in observable terms only. Another prominent instantialist, Clark Glymour, allows you to use the parts of the theory you are attempting to confirm to infer the presence of unobservable instances—a technique he calls "bootstrapping" (Glymour 1980). To take advantage of bootstrapping, however, you must build all of the auxiliaries you need—except, perhaps, those that are known independently to be true, and thus concerning which there is consensus—into the theory, resulting in the confirmation of a theory-auxiliary conglomerate rather than of the theory alone. (Boot-strapping may in any case be too liberal; see Christensen (1990).)

I have no magic bullet to offer. I can only echo the conventional wisdom: use auxiliaries you can rely on. Seek to test auxiliaries independently of your theory. And so on. Just because you cannot have independence from auxiliaries everywhere, however, is no reason to turn it down when you can get it—and you can get it in the causal-empirical inferential step from instance to hypothesis.

5.4 Transmission

I have defined a direct confirmation relation. Do I need a transmission rule such as Hempel's consequence rule? The closest viable equivalent to the consequence rule is something like the following:

> If *e* directly confirms *h*, then *e* also confirms any part of *h*.

The validity of the transmission rule need not be postulated; it follows from the transitivity of parthood (any part of a part of *h* is a part of *h*) and the definition of I-confirmation in terms of saturating causal-empirical support, by the following argument. Suppose that *e* I-confirms *h*. Then *e* causal-empirically supports all parts of *h*. If *g* is a part of *h*, then by the transitivity of parthood, all parts of *g* are parts of *h*, thus *e* causal-empirically supports all parts of *g*. Consequently, *e* I-confirms *g*.

Such a transmission rule is of limited use, however. Suppose that *e* confirms *h*, and that *h* is in turn a part of a theory *t*, but that *e* does not empirically saturate *t*: there is some part of *t* that is not empirically supported by *e*. We would like to know how to regard *t* in the light of *e*. The transmission rule provides no help.

I propose to handle the problem in three steps. First, observe that it is not a formal defect of instantialism if it has nothing to say about the evidential relation between *t* and *e*. After all, instantialism is a theory of I-confirmation, and I-confirmation was never supposed to provide a universal theory of inductive reasoning (see section 4).

Second, it can be argued that science has no principled interest in *t* as a whole until it is saturated by some piece of evidence—since without such evidence it is no more than either a conjunction of independently I-confirmed hypotheses (if each of its parts is I-confirmed by some piece of evidence), or the conjunction of one or more I-confirmed hypotheses with other hypotheses having, as yet, no empirical basis (otherwise).

Third, when practical or decision-theoretic concerns demand that the possibility of *t*'s truth be taken into account, what is relevant is the total inductive support lent by *e* to *h*, hence not the relation of I-confirmation but the relation of B-confirmation, which being universalistic always has something to say about the impact of *e* on *h*. In short, from a practical point of view what matters is the degree to which *e* B-confirms *t*; from a scientifically purist point of view, what matters is what the available evidence I-confirms—so that if *e* does not saturate *t*, then *t* does not yet matter.

6. A Fatal Objection?

Consider a well-known argument due to I. J. Good. Suppose you know that you have just been transported randomly to one of two worlds. Nine-tenths of the first world's birds are ravens, and nine-tenths of these are black. The chance of a random bird's being a black raven in this world is therefore 81%. In the second world, one bird in ten is a raven, and all of these are black. The chance of a random bird's being a black raven in this world is therefore 10%. You open your eyes. The first bird you see is a black raven. This is much more likely in the first world than the second world, so the observation gives you good reason to think that you are in the first, not the second world. In the first world, however, not all ravens are black. Thus your observation of a black raven has given you good reason to disbelieve that all ravens in your new world are black. As Good rightly reasons, the black raven, uncontroversially an instance of the raven hypothesis, provides strong inductive support for the hypothesis's negation (Good 1967).

It is natural to conclude that the black raven disconfirms the raven hypothesis, thus entirely undermining, in a paragraph, the instantialist approach to confirmation, and furthermore, showing that it fails precisely because it neglects epistemic context.

The refutation of instantialism goes through, however, only if is supposed that the sense in which Good's black raven disconfirms the raven hypothesis is identical to the sense in which instantialism holds that all instances confirm their hypotheses. I, of course, deny this: Good's raven B-disconfirms the raven hypothesis, but instantialism is a theory of I-confirmation. To put the point in terms of inductive support, Good's raven bears negatively on the raven hypothesis all things considered, but instantialism makes claims only about path-specific empirical support. The evidence's providing positive support for a hypothesis along one path is quite consistent with its providing sufficient negative support along other paths that the net effect is inductively undermining rather than affirming.[11] (This is not, note, the defense mounted by Hempel (1967). Hempel, I think, has universalistic aspirations for this confirmation relation—a serious error, on my approach.)

A question remains. The proponent of instance confirmation is committed to Good's black raven's lending positive inductive support to the raven hypothesis along at least one path. But the significance of the raven seems to be contained entirely in the information you have about your two worlds, by way of which the raven disconfirms the hypothesis. Where does the putative positive support go?

The answer is that it is "trumped" or completely overridden. Compare the following simple case: h confers a high physical probability on e, but e together with background knowledge entails that h is false. In virtue of its conferring high physical probability on the evidence, the hypothesis is supported by the evidence along one path. But in virtue of entailing, in conjunction with background knowledge, the negation of the hypothesis, the hypothesis is refuted. These two evidential relations are not additive—their net effect is not obtained simply by taking the sum of the parts. Rather, the latter, refuting relation entirely nullifies the former. It is what some inductive logicians would call a "rebutting defeater" (Pollock 1974).

11. More generally, there are two ways to deal Good-style counterexamples, that is, cases where an instance B-disconfirms its hypothesis. First, you might show that the disconfirmation is not empirical, that is, that the negative inductive support that secures the disconfirmation is not of an empirical variety. Second, you might show that the negative inductive support responsible for the disconfirmation of the hypothesis, while empirical, is not along the same path in virtue of which the evidence is an instance of the hypothesis.

It is a challenge to discern a nullified support relation between evidence and hypothesis in a purely all-things-considered inductive system such as Bayesian confirmation theory. One hope is to examine certain counterfactual B-confirmation relations, the relations that would exist if background knowledge were somewhat different. (Compare the similar solutions suggested to the problem of old evidence in, for example, Howson and Urbach (1993).) I regard this, however, as a weakness of Bayesian confirmation theory, rather than as good reason to think that there is no sense to the notion of nullified (or for that matter, path-specific) support.

7. Conclusion

I have argued for the existence of two evidential relations: B-confirmation, which is a relation of universalistic, perhaps all-things-considered, inductive support, and I-confirmation, which is a relation of path-specific empirical inductive saturation. Both relations might be called the confirmation relation, the first perhaps more reasonably in the context of decision theory or some other practical pursuit, the second more reasonably in the narrower context of scientific inquiry. From the existence of this distinction and the supporting arguments, I draw several different kinds of conclusions.

First, not so much a conclusion as a piece of consciousness-raising: a theory of confirmation does not have to be a universalistic theory of inductive support, that is, it does not have to account for every evidential relation in which one proposition gives us good reason to believe or disbelieve another. Confirmation theory is a part of philosophy of science, and a far narrower relation, of special interest to and significance for science, might be taken as its proper area of inquiry—without in any way disparaging the enterprise, pursued by Bayesians and others, of giving a complete theory of inductive reasoning.

Second, a historical conclusion: Hempel, Glymour, and other instantialists should be viewed as providing, not an account of universalistic inductive support or B-confirmation, but an account of a more selective evidential relation, one that is at the very least path specific. Their theories of confirmation make the most sense, I think, when interpreted as accounts of I-confirmation, that is, of a relation of saturating empirical support. This explains their favorable attitude to the consequence rule, and even more important, their focus on instances, which are after all—outside of the lab—a rather exclusive class of inductive supporters.

That said, it must be admitted that Hempel sometimes writes as though he is pursuing a universalistic notion of confirmation. In the introduction to his classic paper on confirmation he says:

> While criteria of valid deduction can be and have been supplied by formal logic, no satisfactory theory providing general criteria of confirmation and disconfirmation appears to be available so far (Hempel 1945, 4).

Hempel's pairing of confirmation and deductive logic in this passage suggests that he envisages a theory of confirmation that is a theory of all inductive inference. Further, as noted in the previous section, he declines to take the non-universalist's obvious way out of Good's problem.

Third, confirmation theory can stand to pay far closer attention to actual scientific method without surrendering its logical and epistemological compass. When real scientists talk about "the evidence", they have in mind facts bearing on their hypotheses in a very particular way, namely, through the relation of empirical support. The nature and importance of empirical support in the scientific enterprise has suffered serious neglect in recent philosophy of science, largely due to a focus on Bayesian universalism. But it is possible, as Hempel and Glymour and section 5 of the present paper show, to investigate empirical support without traducing the precepts of traditional epistemology, that is, without abandoning your commitment to the precise and relatively formal specification of the underpinnings of the relevant evidential relations and without renouncing as your paramount goal an explanation of their inductive force.

Fourth, there is much work to be done in confirmation theory yet. The account of instantiation given in section 5 is just a beginning. The topic of empirical support ought to be revived. Work on empirical support might take the more narrowly epistemological direction described in the last paragraph, asking for example whether there are non-causal forms of empirical support, but it might also set out in a broader methodological direction, asking about the social function of the focus on empirical support in science. The Bayesian ride has been a thrill, but it is time for something new.

Fifth, there is a lesson in all of this to be learned about inference to the best explanation. Laws of nature and other sufficiently exalted generalizations explain their instances; they are also confirmed by their instances. In the case of causal generalizations, the explanation and the confirmation are in virtue of almost the same facts, namely, the instance's having been actually or apparently produced in accordance with an associated causal model. As the actual/apparent dichotomy shows, the parallel is not exact: a false hypothesis can be confirmed by its instances, but it cannot genuinely explain those instances. (A more precise parallel, then, is between confirmation and apparent explanation.) But it is close enough, I think, to account for a part of our sense that much scientific inference flows along a connection that is, when reversed, explanatory.

References

Christensen, D. (1990). The irrelevance of bootstrapping. *Philosophy of Science* 57:644–662.

Fitelson, B. (2002). Putting the irrelevance back into the problem of irrelevant conjunction. *Philosophy of Science* 69:611–622.

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

Good, I. J. (1967). The white shoe is a red herring. *British Journal for the Philosophy of Science* 17:322.

Hempel, C. (1945). Studies in the logic of confirmation. *Mind* 54:1–26, 97–121. Reprinted in *Aspects of Scientific Explanation*, chap. 1. Free Press, New York.

———. (1967). The white shoe: No red herring. *British Journal for the Philosophy of Science* 18:239–240.

Howson, C. and P. Urbach. (1993). *Scientific Reasoning: The Bayesian Approach*. Second edition. Open Court, Chicago.

Kitcher, P. (1993). *The Advancement of Science*. Oxford University Press, Oxford.

Pollock, J. L. (1974). *Knowledge and Justification*. Princeton University Press, Princeton, NJ.

Strevens, M. (2000). The essentialist aspect of naive theories. *Cognition* 74:149–175.

———. (2001). The Bayesian treatment of auxiliary hypotheses. *British Journal for the Philosophy of Science* 52:515–538.

———. (2004). The causal and unification accounts of explanation unified – causally. *Noûs* 38:154–176.

———. (Manuscript a). Ceteris paribus hedges and the role of causal hypotheses in science.

———. (Manuscript b). *Depth: An Account of Scientific Explanation*.

———. (Manuscript c). What are special science laws made of?