Bayesianism versus Confirmation

Michael Strevens

Draft of March 2013

Abstract: The usual Bayesian approach to understanding the confirmation of scientific theories is inadequate. The problem lies not with Bayesian epistemology, but with a simplistic equation of the subjective, individualistic evidential relevance relation that Bayesianism attempts to capture and the more objective relevance relation of confirmation.

1. Grant that Bayesian epistemology provides a respectable theory of rational belief revision; my question in this paper is how the Bayesian theory of belief revision is related to the confirmation relation in science.

The core of a Bayesian account of theory confirmation is typically thought to be straightforward:

> *Simple Proposal:* A piece of evidence confirms a theory for a reasoner just in case the evidence increases the reasoner's subjective probability for the theory (in accordance with the Bayesian conditionalization rule), and it disconfirms a theory just in case the evidence decreases the subjective probability.

As implied by the wording, confirmation is always relative to a reasoner: what increases the subjective probability of a hypothesis for one scientist may decrease it for another.

This bare account of the relation between Bayesian probability kinematics and scientific confirmation can be augmented in various ways. You might introduce a notion of degree of confirmation (Milne 1996; Fitelson 1999), or a more static notion of evidential relevance that allows a piece of evidence to remain relevant to a hypothesis even after it has had its one-time effect on the hypothesis's probability (Garber 1983). These elaborations will not, however, concern me here. I want to argue that the Simple Proposal presented above is incorrect, that it must be replaced with something that better captures the distinctive features of science's confirmation relation.

2. A scientist performs an experiment to test hypothesis h. The scientist's background beliefs, together with h itself, logically entail that the result of the experiment will be e. The scientist, however, incorrectly believes that the theory and background entail $\neg e$. As it happens, the experiment produces result $\neg e$, contrary to the hypothesis's prediction. The scientist, however, thinks that the experiment has produced just the result that the hypothesis foresees. Is the hypothesis confirmed or disconfirmed?

The Bayesian has two ways of thinking about this problem. Bayesian epistemology, in its classical form, incorporates an assumption of "logical omniscience": a Bayesian is supposed to be logically infallible (and is supposed to be aware of all theoretical possibilities). The classical Bayesian model cannot, consequently, be used to represent reasoning in which logical errors are made; it simply does not pronounce on the case at hand.

It is, however, widely recognized among those who favor a Bayesian representation of scientific reasoning that such reasoning does not conform to the assumption of logical omniscience. A more tolerant or "humanized" variant of Bayesianism is needed, on which, for example, logical tautologies can be attributed degrees of belief less than one (Garber 1983; Jeffrey 1983). The technical obstacles to constructing such an account are serious, but put all such worries on hold and suppose that we are dealing with this sort of logically fallibilist Bayesian theory.

A humanized Bayesianism will treat the experimenter as having a high subjective probability for $\neg e$ conditional on h (though the axioms of probability require a zero probability). Consequently, conditionalizing on $\neg e$ will boost the experimenter's subjective probability for h.[1] The humanized Bayesian judges, then, that the experimental result in the above scenario confirms the hypothesis—if the Simple Proposal is correct. (More exactly, it confirms the hypothesis for the scientist in question; it would disconfirm the hypothesis for a bystander with the same background beliefs who correctly believed that the hypothesis predicts e.)

Is this right? It is not. The way that scientists think and talk about the relation between theory and evidence—about the confirmation relation—mandates the following description of the case: the experimental result disconfirms the hypothesis, but the experimenter because of their logical error falsely believes that the result confirms it. What matters for confirmation, then, is the objective logical relation between the hypothesis and the evidence, not what a scientist takes that relation to be.

Before you begin to raise objections, let me describe two other cases that I take to raise similar problems for Bayesian confirmation theory.

Consider a scientist investigating the hypothesis *Gold is diamagnetic*. They test a specimen of what they believe to be gold, and find that it is paramagnetic, and thus not diamagnetic. The sample, however, is not gold but iron pyrite (fool's gold). Because the scientist is confident, and by assumption remains confident, that the sample is gold, their subjective probability for its being paramagnetic conditional on the hypothesis is very low, and certainly, much lower than their unconditional probability for paramagnetism. So their subjective probability for the hypothesis will decrease on their learning of the paramagnetic quality of the sample. According to the Simple Proposal, then, the fact of the sample's paramagnetism disconfirms the hypothesis.

1. I am assuming of course that the experimenter's unconditional probability for $\neg e$ is lower than this conditional probability.

This is, again, incorrect. Because the hypothesis makes no claims about the magnetic properties of iron pyrite, the observation of these properties, whatever they may be, is irrelevant to the hypothesis. The pyrite's paramagnetism neither confirms nor disconfirms the hypothesis, which has things to say only about gold. The investigator falsely believes that the hypothesis is disconfirmed, but they are wrong in this belief.

Or consider the computer program PYTHIA that is used to derive statistical predictions from the complex mathematics of theoretical particle physics. Suppose that an experimenter testing one of these theories—call it Ubersymmetry—uses PYTHIA to infer the physical probability that the theory ascribes to some particular event.[2] They have every reason to trust PYTHIA, as it has been exhaustively tested and extensively used. But the program may not be free of errors. For that matter, a computer on which the program is run may malfunction. In such circumstances, the experimenter might well come to have a false belief about the physical probability that Ubersymmetry assigns to an event. Suppose that this unfortunate train of events comes to pass. As a consequence, the experimenter falsely—though blamelessly—believes that Ubersymmetry assigns a physical probability to an event that is quite at odds with the event's observed frequency, when in fact the frequency and the probability are close.[3]

The experimenter's subjective probability for Ubersymmetry will drop; had they had the correct belief about the physical probability, it would rather have risen. Does the evidence confirm or disconfirm Ubersymmetry? It

2. Because PYTHIA uses Monte Carlo methods, the inference is inductive rather than deductive even if the issue of the fallibility of the program and machine are put aside. This aspect of PYTHIA is irrelevant to the example (though of course it might be used to generate a similar scenario).

3. In this case, then, the experimenter violates Lewis's Principal Principle (Lewis 1980), which like other principles of probability coordination requires (in the kind of circumstances I have in mind here) that the subjective probability for the evidence conditional on the hypothesis equal the physical probability ascribed to the evidence by the hypothesis. The violation is blameless, indeed rational, because the experimenter has good reason to believe that PYTHIA is reliable.

confirms it, I suggest: in scientific discourse, confirmation is matter of what a theory actually predicts, not of what some experimenter thinks it predicts. The Simple Proposal fails to deliver this judgment.

3. What is wrong with the notion of confirmation offered by the conjunction of Bayesian epistemology and the Simple Proposal is that it is in a certain sense too subjective: it implies that a hypothesis is confirmed (for a scientist) whenever the scientist *thinks* that it is confirmed. As the examples above show, however, confirmation as it figures in science is a more objective relation than that. Whether the evidence confirms a hypothesis depends on what the hypothesis says about the evidence, not what some scientist believes—perhaps mistakenly—that the hypothesis says about the evidence.

Where does the fault lie? Not, I suggest, in Bayesian epistemology. In the cases above, Bayesian updating is doing the job it is supposed to do, capturing a sense of evidential relevance that is germane to individualistic epistemology, that is, to deciding the question of what an individual ought, given their epistemic situation, to believe. This is most obvious in the PYTHIA case, where I have given you enough of the epistemic background to make it clear that the scientist in question ought to take PYTHIA's (incorrect) judgments seriously and lower their confidence in the Ubersymmetry theory, even though to do so is to move away from the truth. But the same can be said of the other cases, too: given the scientist's belief that their sample is gold, for example, its paramagnetism provides good reason to relinquish the hypothesis that gold is diamagnetic.

The problem, then, must be attributed to the Simple Proposal, which attempts to capture the relation of scientific confirmation using an account of the individualistic evidential relevance relation. Confirmation is an evidential relevance relation, to be sure, but it is not individualistic in this sense. Science is an intersubjective, indeed a social, pursuit. For this and other reasons (Strevens manuscript), it requires a more objective conception of evidential relevance, a conception on which it is possible for an individual scientist to fail to grasp the evidential significance of an observation.[4]

Two brief comments. First, that Bayesian confirmation theory is "too subjective" is a well-known concern. Such a complaint is insufficiently fine-grained, in my view; it ought to be divided in two. On the one hand, you can criticize the Simple Proposal as too subjective, as I have here, because it attempts to align the more objective confirmation relation with the more subjective relation of evidential relevance proper to an individualistic epistemology. On the other hand, you can criticize Bayesian epistemology as too subjective even as an individualistic epistemology, on the grounds that its relations of evidential relevance are entirely determined by a personal assignment of prior probabilities on which there are almost no constraints. Both objections perhaps have bite, but it is a mistake to assume that what is too subjective for science is thereby also too subjective for the individual.

Second, you might wonder whether scientists use the term "confirmation" to talk about many different evidential support relations, some more subjective and some more objective, in different contexts. Perhaps there is some flexibility in the term, but I think (although in this short paper I have not offered sociological evidence) that there is a special role in scientific discourse for a relation based on the actual rather than the fancied implications of a theory. Even if you do not think that this relation dominates all others in scientific discourse, I hope I will have persuaded you at least that it has an important place, and so that there are significant uses of the term "confirms" that are not captured by the Bayesian conception of support.

4. To move forward, what is needed? Two things: an account of the confirmation relation that is relatively independent of individualistic epistemology, and a replacement for the Simple Proposal that gives a more sophisticated account of the relationship between individualistic epistemology and the confirmation

4. For some other ways in which the confirmation relation is intersubjective or objective rather than individualistic, see Strevens (2009) and Strevens (2010).

relation. I will say no more about the former project (for the beginnings of an account of the confirmation relation, see Strevens (manuscript)). I do have a few concluding, schematic comments about the latter project, that is, about the proper interface between individualistic epistemology and science's confirmation relation.

The individualistic relevance relation cannot be equated with the confirmation relation. But facts about the confirmation relation can and will enter into an individual's reasoning, and thus into individualistic epistemology, by way of beliefs about the confirmation relation. In the PYTHIA case, for example, the drop in the experimenter's subjective probability for the theory reflects their (false) belief that the evidence disconfirms the theory.[5] These beliefs about confirmation will of course closely track beliefs about what the hypothesis says about the evidence, which may in turn track beliefs about the logical implications of the hypothesis (for example, what physical probability the hypothesis assigns to an event), beliefs about the properties of the evidence (for example, whether a specimen is gold), or perhaps even beliefs about the content of a ceteris paribus hedge (Strevens forthcoming). The proper tracking relation might be captured by the following sort of constraint: ceteris paribus, the recognition that a hypothesis is confirmed should provide individualistic good reason to increase your subjective probability for the hypothesis.

A comprehensive individualistic epistemology, such as humanized Bayesianism, will incorporate all the forms of uncertainty described in the previous paragraph. The good news, then, is that such an epistemology should be able to interface with any reasonable theory of the nature of confirmation. The corollary of this flexibility—the bad news, for those who like things simple—is that an adequate theory of the confirmation relation cannot be read off from the correct theory of individualistic epistemology, whether it is Bayesianism

5. The Bayesian apparatus, incidentally, provides no resources for saying that one of these facts is grounded in the other.

or something else.

References

Earman, J. (ed.). (1983). *Testing Scientific Theories*, volume 10 of *Minnesota Studies in the Philosophy of Science*. University of Minnesota Press, Minneapolis.

Fitelson, B. (1999). The plurality of Bayesian measures of confirmation and the problem of measure sensitivity. *Philosophy of Science* 66:S362–S378.

Garber, D. (1983). Old evidence and logical omniscience in Bayesian confirmation theory. In Earman (1983).

Jeffrey, R. C. (1983). Bayesianism with a human face. In Earman (1983).

Lewis, D. (1980). A subjectivist's guide to objective chance. In R. C. Jeffrey (ed.), *Studies in Inductive Logic and Probability*, volume 2. University of California Press, Berkeley, CA.

Milne, P. (1996). $\log\left[P(h|eb)/P(h|b)\right]$ is the one true measure of confirmation. *Philosophy of Science* 63:21–26.

Strevens, M. (2009). Objective evidence and absence. *Philosophical Studies* 143:91–100.

———. (2010). Reconsidering authority: Scientific expertise, bounded rationality, and epistemic backtracking. *Oxford Studies in Epistemology* 3:294–330.

———. (Forthcoming). Ceteris paribus hedges: Causal voodoo that works. *Journal of Philosophy*.

———. (Manuscript). What is empirical testing?