Smith on Justification and Probability

Brian Weatherson

1 Models of Justification

Call *Justificatory Probabilism* (hereafter, JP) the thesis that there is some (classical) probability function Pr such that for an agent S with evidence E, the degree to which they are justified in believing a hypothesis H is given by $\Pr(H|E)$. As stated, the thesis is fairly ambiguous, though none of the disambiguations are obviously true. Indeed, several of them are obviously false.

If JP is a thesis about how justified agents are in fully believing propositions, it is trivially false. I'm about to flip a penny. Call H the proposition that it will land heads. Right now I'm completely unjustified in believing either H or $\neg H$. Yet according to JP, at least one of them must be half-justified.

Richard Holton (2008) has argued that full belief comes in degrees. This is an attractive thesis, especially on a pragmatic view of belief. Start with the picture that beliefs are, as Ross and Schroeder (2014) say, something like those propositions we are disposed to take as given in inquiry. Now in different inquiries we will take different things as simply given (Hawthorne, 2004). So we might measure the strength of a full belief as something like the range of inquiries in which we'll take it as given. Then we could interpret JP as a thesis about how strongly S is justified in believing H. But again, it would be false, for the same reason as given in the previous paragraph. We are completely unjustified in taking as given either that a coin will land heads or that it will land tails, contra JP.

It is more plausible to take JP as a thesis about *credences*, one that has consequences for the theory of belief given some connection between beliefs and credences. What thesis could it be? The following three candidates spring to mind.

- The one and only credence which S is justified in having in H given evidence E is $\Pr(H|E)$.
- The highest credence which S is justified in having in H given evidence E is $\Pr(H|E)$.
- The lowest credence which S is justified in having in H given evidence E is $\Pr(H|E)$.

But none of these theses can be true unless there is a unique credence which S is justified in having in H with evidence E. Perhaps that's true, it has been defended by Roger White (2005) for example. But it is a strong and I think rather unintuitive thesis.

So it is hard to state a plausible version of JP. The only plausible version assumes a strong uniqueness thesis, and isn't even in the first instance a thesis about beliefs.

† Unpublished. Thanks to participants in grad seminars at Rutgers and Michigan for very useful feedback.

It is a thesis about credences, and needs some supplementing with a theory about the belief-credence link to tell us *anything* about norms of belief.

But for all that JP has one signal advantage. Having a formal model to test our informal and intuitive ideas against is often crucial in making progress, and stopping debates from turning into clashes of intuitions. And indeed, whatever can be said about a priori, foundational arguments for JP, it has I think been very *useful* in performing just this role in epistemology of the last few decades.[1] But there's no reason to think we couldn't come up with *other* formal models that could be similarly useful in providing a formal test of epistemological theories.

In an excellent recent paper in *Noûs*, Martin Smith (2010) does just this. The purpose of this note is to extend Smith's model in a way that makes it even less like the probabilistic models he aims to offer an alternative to.

2 Conjunctions and Justification

The particular feature of JP that Smith wants to escape concerns the way it handles conjunctions. At least given most treatments of the link between credences and beliefs, JP has the following feature. Some propositions can each have an arbitrarily high justification short of maximal justification, while their conjunction has an arbitrarily low justification, assuming the propositions in question are allowed to be sufficiently numerous. That follows (given a credence-belief link of a suitable kind) from the fact that some propositions can each have arbitrarily high probability short of 1, while their conjunction has arbitrarily low probability. Some authors, most notably David Christensen (2005) have defended this consequence of JP. Indeed, Christensen says that reflections on the preface paradox show that it is a feature, not a bug, of the model. But many others have found it strange that an agent who is justified in believing A to some degree, and justified in believing B to the same degree, may yet need more evidence to be justified in believing $A \wedge B$ to the same degree.

Smith's alternative model doesn't have that feature. Here's how his model works. Assign to each world a 'normalcy ranking'. Smith doesn't fill in the details in exactly the way I'm about to, but I don't think it loses anything to do it this way. The normalcy rankings are non-negative integers, with the most normal worlds getting a rank of 0, and higher ranks meaning that the worlds are less normal. Then given that her evidence is E, the degree to which S is justified in believing H is the smallest n such that there is a world in which $E \wedge \neg H$ with normalcy rank n. If there is no such world, i.e., if E entails H, then the agent's degree of justification is ∞.[2] For notational convenience, we'll write $J(H, E)$ for the degree of justification of H by

[1] Alan Hájek (2008) has a much more thorough discussion of the attempts to provide such an a priori, foundational, defence of JP.

[2] There is obviously some heavy idealisation going on in here. Every logical and mathematical truth gets a justification of ∞ for every agent. If we take worlds to be metaphysically possible worlds, then every necessary truth gets a justification of ∞ for every agent. That's obviously crazy taken literally. The response is that, like the probabilistic model of justification, this is just a model. Finding a mathematical model for how justified mathematically limited creatures are in their mathematical beliefs is, to put it mildly, hard. We shouldn't put off the work of building formal models until we solve that hard problem, though we shouldn't ignore that it is a worthwhile problem to solve either.

evidence E in what follows.

There are already some interesting topological differences between this model of justification and the probabilistic model. Notably, the degrees of justification that the model issues are not *dense*. That's probably a good thing; it feels like an artifact of the probabilistic model that it does issue in densely packed degrees of support. Actual justification feels more coarse-grained than that.

But a bigger difference is in how Smith's model treats justification of conjunctions. Consider the level of justification an agent has in $A \wedge B$. That will be lowest normalcy rank that attaches to a world in which $E \wedge \neg(A \wedge B)$. At that world, either A or B will be false. And of course E will be true. So it will either be a world in which $E \wedge \neg A$, or in which $E \wedge \neg B$. So it can't be that both A and B are more justified than $A \wedge B$. Actually, it isn't too hard to prove that $J(A \wedge B, E)$ equals the lower of $J(A, E)$ and $J(B, E)$. (Or both, if they are equal.) To many of us, that's a nice intuitive feature of the model.

3 Conditionals and Justification

One of the nice things about the probabilistic model of justification is that it doesn't just tell us a story about justification at a time (i.e., a *synchronic* story about justification), it also gives us a dynamic, diachronic story about how justification evolves over time. When you add evidence E, the new justification for H equals the old justification for H given E. It is easy enough to prove the following well-known theorem.

$$Pr(H|E) \leq Pr(E \supset H)$$

That suggests that an agent can never become more justified in believing H once they get E than they could have antecedently been in $E \supset H$. And on reflection, that's a slightly puzzling feature of JP. Consider the extreme case, where E is all the empirical evidence the agent has, and H is something well supported, but not entailed, by E. The model suggests that she can't be more justified in believing H right now than she was in believing $E \supset H$ before she got E. But before she got E, she was reasoning *a priori*. And since E doesn't entail H there are possible worlds in which $E \supset H$ is false. One might think that, *a priori* at least, there is no reason to distinguish between possible propositions.[3] That is, *a priori* she has no justification for believing $E \supset H$ at all.

The considerations of the last paragraph should be familiar from some recent debates concerning dogmatism. Roger White (2006) has used a similar argument to show that dogmatism, in the sense of James Pryor (2000) is incompatible with JP. I'm inclined to think that's more of a problem for JP than for dogmatism, but that's for another day. What I want to note here is that very simple probabilistic reasoning gets us a strong, and perhaps implausible, claim about *a priori* justification.

---

[3]If the worlds are metaphysically possible worlds, this won't quite be right. After all, there are contingent *a priori* truths. I'm assuming the worlds here are something like the epistemic possibilities of Chalmers (2011). From now on, all references to worlds in what follows should be assumed to be references to these kinds of epistemic possibilities; these are what really matter to epistemology.

We should investigate whether this is a real strength of the probabilistic model of justification, or an unfortunate artefact of that model.

Surprisingly, the 'normalcy' story that Martin Smith tells has a similar consequence. Note that we can prove the following theorem.

$$J(H, K \wedge E) = J(E \supset H, K)$$

Think of the K here as a 'background' evidence. If we are reasoning *a priori*, we can let K be the conjunction of all *a priori* knowable truths. What the claim says is that how justified H is by an evidence set to which E is added is exactly the same as how justified $E \supset H$ is by that evidence set. The proof of this is pretty easy. By definition, $J(H, K \wedge E)$ is the normalcy of the most normal world in $K \wedge E \wedge \neg H$. And $J(E \supset H, K)$ is the normalcy of the most normal world in $K \wedge \neg(E \supset H)$. But $\neg(E \supset H)$ is true iff $E \wedge \neg H$ is true. So $K \wedge E \wedge \neg H$ is the exact same set as $K \wedge \neg(E \supset H)$, hence $J(H, K \wedge E) = J(E \supset H, K)$, as required.

We can turn this into a story about justificatory dynamics with a common simplifying assumption. Say that $J_E(H, K)$ is how justified is for an agent who antecedently had evidence K once she learns that E. The simplifying assumption is that $J_E(H, K) = J(H, K \wedge E)$. In general this probably isn't right; sometimes E defeats something that's previously known and hence removes it from the agent's evidence. With the simplifying assumption, we can prove that $J_E(H, K) = J(E \supset H, K)$. If E defeats some previous evidence, we'll get a slightly weaker result (whose proof I omit here), namely that $J_E(H, K) \leq J(E \supset H, K)$. But what we can never do is acquire evidence E, and hence become more justified in believing H than we antecedently were in believing $E \supset H$.

4 Making Normalcy Evidence-Relative

It seems to me that this is a weakness of Smith's model, a weakness it shares with the probabilistic model. Fortunately, there is a small tweak that avoids this weakness. And the tweak helps us explain what normalcy might be in a philosophically revealing way.

So far we've talked about *the* normalcy of a world. Let's drop the assumption that there is any such thing. Actually, that's probably a good assumption to drop. Is this world normal? Well, it seems so to us, but from some very different perspectives, it probably looks like a place where weird and wonderful things happen all the time in very abnormal ways. Instead of saying that there is such a thing as the normalcy of a world, we'll say that normalcy is relative to evidence.

To introduce some terminology, say $N(w, E)$ is the normalcy of world w given evidence E. We'll then say that $J(H, E)$ is the lowest value of $N(w, E)$ for a w where $E \wedge \neg H$. Note that evidence is now playing two roles. In Smith's model, it plays just the one role; it rules out incompatible worlds. The level of justification for H is then the lowest normalcy rank of a remaining world in which $\neg H$. In this model, it plays two roles. First it determines which worlds are more and less normal. Then it eliminates incompatible worlds. And again the level of justification for H is then the lowest normalcy rank of a remaining world in which $\neg H$.

What could determine $N(w, E)$? I think some kind of similarity metric could work. Some worlds resemble our evidence. Before a coin is flipped, a world in which it lands tails and a world in which it lands heads both resemble our evidence. Our evidence, after all, includes similar coins landing heads and landing tails. But a world in which it snows in Miami next August does not resemble our evidence, which hasn't included snow at any similar locale for many years.

We have to be a little careful here because of limitations of our evidence. If it is now t, then on a broadly causal model of evidence, all of our evidence will concern facts about times before t. But worlds that extend beyond t are not thereby non-normal. The thing to do here is to restrict the aspects of similarity that matter. In particular, we should ignore respects of similarity or difference with respect to what Sider (2001) calls *maximal* properties. But once that's taken into account, we get a nice picture of what makes worlds normal for an agent; they are worlds where things are like what the agent has perceived.[4]

This picture suggests a restriction on the N function. If the agent has no (empirical) evidence, then any world resembles the agent's evidence just as well as any other. So $N(w, \emptyset) = 0$ for all w. And this implies that anything which is false at any world is completely unjustified *a priori*. But it doesn't follow that only things entailed by one's evidence are justified. Returning to an example from above, let E be the evidence that you actually have, and H be that it won't snow in Miami next August. The most normal worlds in which your evidence is true and it does snow in Miami next August are still pretty weird. That is, those worlds look nothing like your evidence. For any such world w, $N(w)$ is huge. So it follows that H is justified for you to a very high degree, even though *a priori* you had no justification to believe $E \supset H$.

5 Summing Up

The point of this note is not that there's something wrong with the probabilistic model of justification, or Smith's alternative model based on normalcy functions. In a way, that would be too easy to prove. After all, both models make it too easy to be justified in believing mathematical truths. But harping on that point would be to ignore the value to researchers of good models, and how models could be good while still delivering absurd results in some class of cases.

Still, we best not ignore that models have limitations. Simply inferring from the probabilistic model that it is possible to be more justified in believing both A and B than one is in believing $A \wedge B$ would be absurd; it would be to take one potentially quirky feature of the model as revealing a deep truth about justification. What we need is an argument that this is a good feature of the model. As I noted earlier, that's what Christensen (2005) does, and his arguments show us a lot about the strengths (and weaknesses) of the probabilistic model. One of the ways in which we test whether an aspect of a model is revealing of an underlying truth, or simply a quirk of the model, is whether it is shared by other natural ways of modelling the phenomenon. One important strength of Smith's work is that it shows us how easy

[4]Thanks here to Neil Mehta.

it is to build a model which doesn't mirror the probabilistic model in this respect, and that makes it more plausible that this is a bug not a real feature of the probabilistic model.

By making a small extension to Smith's model, I've attempted to do the same thing for the claim that learning E never makes H more justified than $E \supset H$ was before learning E. Some reflections on *a priori* justification suggest that this shouldn't be a universal truth, but on both Smith's model and the probabilistic model it is always the case. But models have limits, and seeing how easy it is to remove this feature of the model, by simply making normalcy relative to what evidence one has, backs up our philosophical intuition that this wasn't a respect in which the model reflected reality.

Further, it is easier to say what evidence relative normalcy really is, i.e., it is similarity in respect of non-maximal properties, than it is to say what it is for a world to be normal in some absolute sense. So my variant on Smith's model uses, perhaps, more understandable primitives. Having said that, both kinds of normalcy (absolute and evidence-relative) seem easier to understand than the magical probability function at the center of JP. But this is all just to say that different models have different strengths and weaknesses. We shouldn't take any one of them to simply prove that something or other is true of real world justification.

That is, the argument that it is possible to be more justified in believing H after getting E than one was in believing $E \supset H$ before getting E is most emphatically *not* that I've created a wonderful model of justification and we can infer that this is possible from the model. Rather, I think we have independent evidence to believe the possibility claim. The purpose of this paper is to show that while this possibility claim can't be modelled in a probabilistic model of justification, it is easy enough to construct a natural model in which it can be modelled. And that eliminates one argument for not believing the possibility claim.

References

Chalmers, David, (2011). "The Nature of Epistemic Space." In Andy Egan and Brian Weatherson (eds.), *Epistemic Modality*, 60-107. Oxford: Oxford University Press. (3)

Christensen, David, (2005). *Putting Logic in Its Place*. Oxford: Oxford University Press. (2, 5)

Hájek, Alan, (2008). "Arguments for-or against-Probabilism?" *British Journal for the Philosophy of Science* 59: 793-819, doi:10.1093/bjps/axn045. (2)

Hawthorne, John, (2004). *Knowledge and Lotteries*. Oxford: Oxford University Press. (1)

Holton, Richard, (2008). "Partial Belief, Partial Intention." *Mind* 117: 27-58, doi:10.1093/mind/fzn002. (1)

Pryor, James, (2000). "The Sceptic and the Dogmatist." *Noûs* 34: 517-549, doi:10.1111/0029-4624.00277. (3)

Ross, Jacob and Schroeder, Mark, (2014). "Belief, Credence, and Pragmatic Encroachment." *Philosophy and Phenomenological Research* 88: 259-288, doi:10.1111/j.1933-1592.2011.00552.x. (1)

Sider, Theodore, (2001). "Maximality and Intrinsic Properties." *Philosophy and Phenomenological Research* 63: 357-364, doi:10.1111/j.1933-1592.2001.tb00109.x. (5)

Smith, Martin, (2010). "What Else Justification Could Be." *Noûs* 44: 10-31, doi:10.1111/j.1468-0068.2009.00729.x. (2)

White, Roger, (2005). "Epistemic permissiveness." *Philosophical Perspectives* 19: 445-459, doi:10.1111/j.1520-8583.2005.00069.x. (1)

—, (2006). "Problems for Dogmatism." *Philosophical Studies* 131: 525-557, doi:10.1007/s11098-004-7487-9. (3)