A non-pragmatic dominance argument for conditionalization

J. Robert G. Williams[*]

June 20, 2013

I'm not omniscient. Even if my beliefs are perfectly rational, they're not perfectly accurate. Given the limited information at my disposal, middling confidence may be a rationally impeccable attitude to have to the proposition *the number of cats in the house is even*. Yet if there are in fact two cats in the house, the proposition is true and the maximally accurate attitude would be full confidence. Or so asserts an influential characterization (Joyce's) of the 'accuracy' of partial beliefs.[1]

Nevertheless if I'm to claim to be perfectly rational, I better avoid *gratuitous* inaccuracy—inaccuracy where my limited information is no excuse. If an agent has belief state b, and belief state b' is more accurate than b no matter what the world is like, then b is rationally flawed. Flawless belief cannot be 'accuracy dominated'. Or so Joyce supposes.

This normative assumption relating accuracy to rationality becomes a premise in an argument for *probabilism* when conjoined to the following theorem: that if one's belief state fails to be probabilistic, it will be gratuitously inaccurate in just the way flagged above.

This note examines the principles required for this argument to go through, and argues that they motivate a (non-pragmatic) argument for Conditionalization as the rationally required update-rule.[2]

Godlike belief

At the centre of this argument for probabilism is the *comparative accuracy* of belief states at worlds. The notion is appealed to in the criterion for rationally flawed belief in the second paragraph. Various assumptions about the features of this 'inaccuracy score' are required to prove the theorem cited in the third paragraph. One candidate 'accuracy scoring rule' is the *Brier score*, which tells us that the degree of inaccuracy of b at w is given by the square difference between the truth value of a proposition at w and the degree of belief b invests in it, summed over all propositions. Other scoring rules are available, but the differences won't matter here.[3]

I want to flag one central feature of the Brier score (and indeed, all the other scoring rules that I've seen). The 'truth value' of a proposition is represented by a numerical value—1 for truth, 0 for falsity. That's clearly crucial to making the proposal well-formed. One cannot perform arithmetical manipulations on 'the True' or 'the False'—only on numbers.[4]

---

[1](Joyce, 1998, 2009)

[2]Throughout, I will be confining attention to belief states over a finite set of worlds. For discussion of extensions to the infinite case, see (Easwaran, 2013).

[3]The results which will be appealed to below go through for any accuracy measure meeting the conditions that (Joyce, 1998) sets out. See (Williams, 2012b) for discussion of the required generalization.

[4]Stephan Hartman, for one, has questioned whether numerical representations of truth values can be justified.

Where do the 1's and 0's come from? Well, there's a belief state that one would *ideally* adopt, given the information that w is actual. We'll call that the 'godlike' credence to have at w. Certainty in A is the godlike attitude to take to A when it is true; utter rejection is the godlike attitude to take to A when it is false. We standardly represent degrees of belief on a scale between 1 and 0, with 1 representing certainty and 0 rejection. So that's where the numbers come from—-they encode information about the godlike credences associated with truth and falsity respectively. As a result, the distribution of truth *values* over propositions at world w exactly coincides with the godlike credences to have at w.

(Association of 1 with truth and 0 with falsity is so widely entrenched that it's easy to let it pass by without comment, and imagine somehow it's true by construction or convention. To appreciate the need for some kind of bridge principle here, it's helpful to imagine for a moment worlds as Lewis (1986) did—maximal spatio-temporally unified chunks of concrete reality. No function from propositions to numbers is intrinsic to worlds so construed! We can also see that a bridge is needed by considering less familiar cases of the same phenomenon. If we have three or more truth statuses to deal with—truth, falsity and 'other'—then it's more obviously a nontrivial question what 'truth value' or 'ideal opinion' we should associate with the new status.[5])

Godlike credence is a normative notion—-it's unpacked as the *ideal* belief state to have given certain information. In contrast to the graded 'degree of accuracy' of an arbitrary belief state at a given world, ideal belief (relative to some information) is absolute, all-or-nothing. You're either ideal or you're not. The picture I want to encourage views the degree of (in)accuracy of b at w as the upshot of two more fundamental facts. The first is that a certain b_w is godlike at w, i.e. that it is the correct belief-state to have, given the information that w is actual. The second is b is such-and-such distance from b_w.

Angelic belief

We will extend the notion of ideal belief, relative to given information. For *any* block of information E, we can ask what one should *ideally* believe, given E as total evidence.

Now, perhaps the question as formulated at the moment doesn't have a categorical answer. Perhaps non-evidential factors—one's attitude to epistemic risk; one's ultimate priors; various culturally entrenched norms, or whatever—deliver different answers. But I'll assume that once we hold fixed all such factors (collectively, call that a *stance*), there is a definite answer to what one should ideally believe, given the evidence. For simplicity, I'll drop explicit mention of the relativization to stances from now on, but the whole discussion to follow will go through relative to a fixed stance.

In the special case where E is complete—where is is the information that w is actual, for some possible world w—then the ideal belief state has a familiar form (at least assuming classicism)—a string of 1s and 0s matching truths and falsities respectively. But if E is less specific, then it's a matter for substantive epistemological debate what shape the ideal belief state, given information E, should be. I'll make no direct assumptions about what these ideal belief states are. What I will do is outline certain normative roles that they play. As we'll see, indirectly, this constrains what sort of thing ideal belief can be.

Whereas ideal belief states relative to complete information are godlike, I'll call the ideal belief states relative to incomplete information *angelic*—still ideal, but lacking omniscience.

Recall our earlier gloss on the normative premise in the dominance argument for probabilism:

---

[5]The last question is addressed in (Williams, 2012b, forthcoming)

the rational agent avoids gratuitous inaccuracy—inaccuracy where our limited information is no excuse. If belief state b' is more accurate than b no matter what the world is like, then b is rationally flawed..

In the light of the discussion in the previous section, we should unpack this as follows:

the rational agent avoids gratuitous distance from the ideal belief state to have at the actual world; that is, she avoids divergence from the ideal where her limited information is no excuse. If belief state b' is closer than b to the *ideal* beliefs to have at the actual world no matter what the world is like, then b is rationally flawed.

I propose we formulate the underlying principle as follows:

Suppose an agent's total evidence (at a time) is E, and her belief state b. Let Γ partition the worlds compatible with E. And let b_γ be the ideal belief state to have relative to total information γ, for each $\gamma \in \Gamma$. Then b is rationally flawed if there is a b' which is closer than b to each b_γ.

The idea is that because Γ partitions the way that the world might be (relative to the evidence possessed by the agent in question), we know that exactly one $\gamma \in \Gamma$ is correct. And consequently, exactly one b_γ reports an ideal belief state relative to information *at least as strong as* the total evidence the agent possesses. Limited information can excuse keeping one's doxastic distance from any single belief state associated with a cell in the partition; but being unnecessarily far away from all of them simply guarantees that one is not as close as one could be to the ideal beliefs one would have in a better-informed state.

Now, for the purposes of the original accuracy argument, we only need the instance of this principle for a single, special partition of propositions—the partition of the *worlds* themselves. But if there's something bad about gratuitously keeping one's doxastic difference from better-informed belief in that case, I can't see why there wouldn't be a similar complaint about gratuitously keeping one's doxastic distance from better-informed *even if not fully informed* belief. So I'm going to endorse this principle in full generality, i.e. relative to an arbitrary partition Γ of the E-worlds.

With that generalization in place, the die is cast—we can start drawing conclusions about the shape of rational belief in general, and ideal belief in particular.

To begin with, we have the Joyce accuracy-domination theorem. Under suitable assumptions about the behaviour of the accuracy measure, one can prove the theorem that an improbabilistic belief state b (one which is not a convex combinations of godlike credences) is always 'accuracy dominated'. We always find a belief state that is guaranteed to be closer to the godlike credence at the actual world than b itself is. And applying the normative principle above in ungeneralized form, relative to the partition of worlds, we conclude that improbabilistic beliefs are rationally flawed.

Given the generalized version, an analogous result holds for other partitions. We can similarly prove that any belief state b that is not a convex combinations of angelic credences (across an arbitrary partition Γ), will always be accompanied by a belief state that is closer than b to each ideal, angelic credence. Hence, the belief state is flawed.[6]

A special case of interest is the following. Take an agent with evidence E, and consider the trivial partition of E—E itself. The angelic belief for that partition is simply b_E—the ideal

---

[6]For this result, relative to the Brier score see (De Finetti, 1974; Williams, 2012a), and for other scoring rules, see (Williams, 2012b).

belief to have for total evidence E. An agent's belief will be gratuitously far away from the ideal, unless they coincide with b_E itself. So rationally flawless beliefs must be always be the ideal response to the total evidence available.

A corollary. Assuming that there are rationally flawless beliefs to have in a situation where the agent has total information E, by the above two observations they must (a) be probabilistic; and (b) be the ideal response to evidence E. It follows that the ideal response to any evidence must also be probabilistic (else the demands placed on a rational belief state given evidence E would be incompatible). I said earlier that we would not *directly assume* anything substantive about the ideal beliefs to have given this or that evidence. And indeed, I did not assume from the get-go that ideal beliefs were probabilistic; we instead *derived* this from the generalized normative role we've given to them.

Update

So far we've being talking about ways of evaluating (categorical) credences. We now move to consider rules for updating belief. I'll assume that the agent's doxastic state includes not only an assignment of (categorical) degrees of belief to propositions, but also an update rule R.

What is an update rule? It is a rule (or function) that takes a belief state, and new evidence, to a new belief state—one that has digested the new information. We expect the update rule to be defined over all well-behaved belief states, though it's no problem if it doesn't tell us what to do for crazy irrational starting points. And we expect it to be defined whenever we gain an increment of evidence, though rules may be silent about cases of non-cumulative evidential change, for example, those involving memory loss. An example is the familiar rule of conditionalization—that for a given (probabilistic) b, and (increment of) new evidence E, we should move to the belief state obtained by conditionalizing b on E in the familiar way. Conditionalization illustrates both kinds of restriction mentioned above: as standardly interpreted it isn't defined for non-probabilistic starting states; nor for cases where we lose evidence rather than accumulating it. This section will argue that an agent with flawless beliefs in the sense spelled out above, must update by Conditionalization.

There is a constraint on an update rule implicit in the constraint of the last section (assuming, as is plausible, that update preserves rational flawlessness). For suppose one starts with a rationally flawless belief state b in a situation where one's total evidence is E_1. And suppose one gets new evidence E (without losing any old) so that one's total new evidence is $E_2 = E_1 \cup E$. We already know one characterization of what a rationally flawless belief state must be like, given the new information state our agent is in. It must be the angelic belief state, relative to E_2. The constraint that arises is that if the update rule R maps b to b^+ given increment of evidence E, and b is the angelic credence for some E_1, then b^+ must be the angelic credence relative to $E_1 \cup E$.

Godlike and angelic belief states now perform two normative functions. Synchronically, they are the base from which we measure the divergence of a given categorical belief state from the ideal—a 'external' evaluation that is applied to agents not in possession of the information that characterizes the ideal (which world is actual, which cell of a partition obtains). Diachronically, they impose a constraint on how we should update our beliefs. The diachronic constraint looks pretty useless right now, given that we still haven't got much information about what the angelic belief states are. But I'll argue below that the only update rule that meets the constraint just articulated, is conditionalization.

The angelic domination result of the previous section tells us that, relative to a partition Γ, an agent's belief state b, if it is to be unflawed, must be a convex combination of the angelic credences corresponding to members of that partition. Now suppose that the agent follows an

update rule R that is normatively ok. A consequence: the angelic credences for information γ, must coincide with $R(b,\gamma)$, for each $\gamma \in \Gamma$. Putting the two sides together: if the agent's current belief state is unflawed and her update rule normatively ok, then her current categorical beliefs have to be a convex combination of the belief states she would update to, upon learning γ, across the various γ that are cells in the partition Γ.

Note what's happened: the mention of angelic belief states drops out altogether, and we're left with an *internal* constraint between her current beliefs and the update rule she follows. The constraint (modulo nuances of philosophical interpretation) is van Fraassen's 'general reflection'. All that's left is to work out the details of what this constraint requires. As it turns out (and here we may appeal to technical results that have been established elsewhere) an update rule R meeting this condition must be conditionalization.[7]

Conclusion

Synchronic probabilism can be supported by accuracy-based arguments. But when we look at the details, we see that the key principle turns on gratuitous divergence between the belief state in question and other, *ideal* belief states. And this motivates a generalization of the basic principle that drives the argument, to states that are ideal relative to partial information. The extended principle enables us to argue for more than just probabilism, however. It establishes a direct link between evidence and rational belief. And it enables us to argue we should update by conditionalization. But this should be qualified in at least three ways.

First, the argument only covers cases over evidence change where no evidence is lost. I take it that this is a qualification that fans of conditionalization need anyway, on pain of declaring amnesia irrational.

Second, the argument does not tell us that conditionalization is always and everywhere the appropriate way to update—even for agents who have probabilistic beliefs. What it tells us is that if your update rule is not conditionalization, then *either* there's something wrong with your categorical degrees of belief, *or* the way you update. Which gets the blame in any given instance is something the argument is silent about. But still: every unflawed doxastic state incorporates conditionalization as its update rule.[8]

---

[7]Here's a sketch of van Fraassen's argument to this effect (van Fraassen, 1999). We have that an agents categorical beliefs b are a convex combination of the updated beliefs relative to γ_i where the γ_i form a partition. So there are λ_i such that $b = \sum_i \lambda_i R(b, \gamma_i)$. We assume that the belief state $R(b, \gamma_i)$ will be credence 1 in any proposition entailed by γ_i (including γ_i itself); and it will be credence 0 in any γ_j for $i \neq j$. After all, this is the belief state one moves to after learning that γ_i, which is compatible with the other γ_j. It follows that $b(\gamma_k) = \sum_i \lambda_i \delta_{ik} = \lambda_k$. Substituting back in, we have that $b = \sum_i b(\gamma_i) R(b, \gamma_i)$—i.e. one's categorical beliefs are the expectation of posterior beliefs upon learning the elements of the partition. But further, we know by the law of total probability that $b(\cdot) = \sum_i b(\gamma_i) b(\cdot | \gamma_i)$. $R(b, \gamma_k)$ and $b(\cdot | \gamma_k)$ are both zero on every proposition incompatible with γ_k. So if they diverge, then they must diverge on some proposition entailed by γ_k. But for all such A, we have $\sum_i b(\gamma_i) R(b, \gamma_i)(A) = b(A) = \sum_i b(\gamma_i) b(A | \gamma_i)$. Both sums are zero everywhere except at k, by the choice of A. And so we have $b(\gamma_k) R(b, \gamma_k)(A) = b(\gamma_k) b(A | \gamma_k)$. Cancelling the $b(\gamma_k)$, this shows that we cannot have any such divergence. So the update must be by conditionalization.

Compare (Skyrms, 2006) and the diachronic dutch book argument in (Williams, 2012a). This note fulfills a promise made in the last citation, to give a non-pragmatic argument for conditionalization, that parallels (and exploits the same geometrical situation as) the pragmatic argument for conditionalization presented therein. The geometrical character of the argument is especially nice, since it allows us to use this result in settings beyond that of standard, classical probabilism. However, it's crucial to note the role played in van Fraassen's argument of the assumption that $R(b, \gamma_i)(\gamma_i) = 1$. This is very plausible, initially. However, in some nonclassical settings, it may fail for the notion of conditional probability required for the rest of the argument (see (Milne, 2008; Williams, forthcoming)).

[8]It's worth noting too that the argument is silent about certain cases that conditionalization traditionally covers. Take a probabilistic belief state b, which is nevertheless desparately out of line with with the agent's evidence (the

Third, there is the concession I made early on—that the ideal belief states the argument appeals to may be relative to one's 'stance' (for example, one's most deep-rooted prejudices). The argument goes through, provided we relativize all appeals to belief to the same underlying stance. We get the result that *holding a fixed stance*, one should update by conditionalization. But here we find a loophole in the argument for conditionalization. Non-conditionalization updates may yet be rationally permissible, if the new information prompts a change of stance. But that's fine: stance-constant conditionalization is conditionalization enough for me.[9]

If successful, this argument supports a genuinely *diachronic* conclusion. It targets and evaluates an update rule directly. This contrasts with a more familiar way of linking conditionalization with accuracy evaluations—via the result that conditionalization maximizes expected accuracy (cf. e.g. (Greaves & Wallace, 2006; Leitgeb & Pettigrew, 2010b,a)). That line of argument seems at best to result in a rational requirement to *plan* to conditionalize. It doesn't tell us that we should *actually* conditionalize, once the information comes in; for plans we've made in the past are not in general rationally binding on us now, if they were based on now-outdated information.[10]

References

De Finetti, Bruno. 1974. *Theory of probability: vol.1*. New York: Wiley.

Easwaran, Kenny. 2013. Expected accuracy supports conditionalization—and conglomerability and reflection. *Philosophy of science*, **80**(1), 119–142.

Greaves, H., & Wallace, D. 2006. Justifying Conditionalization: Conditionalization maximizes expected epistemic utility. *Mind*, **115**(459).

Joyce, James M. 1998. 'A non-pragmatic vindication of probabilism'. *Philosophy of science*, **65**, 575–603.

Joyce, James M. 2009. 'Accuracy and coherence: prospects for an alethic epistemology of partial belief'. *Pages 263–297 of:* Huber, Franz, & Schmidt-Petri, Christoph (eds), *Degrees of belief*. Springer.

Leitgeb, H., & Pettigrew, R. 2010a. 'An objective justification of bayesianism i: Measuring inaccuracy'. *Philosophy of science*, **77**(2), 201–235.

Leitgeb, H., & Pettigrew, R. 2010b. 'An objective justification of bayesianism ii: The consequences of minimizing inaccuracy'. *Philosophy of science*, **77**(2), 236–272.

Lewis, David K. 1986. *On the plurality of worlds*. Oxford: Blackwell.

agent believes that there something square in front of him, despite the manifest presence of a sphere). This belief state is by the arguments of the Angelic Belief section flawed. But still, if the agent now receives new information (he's told Eagles are coming) then conditionalization still tells him how to update. But our principle falls silent. The constraint on update that I gave above assumed that update preserves flawlessness—so that given we start flawlessly, we better end in the (unique) flawless state picked out by enhanced information. But the antecedent fails here. I don't see this as a significant limitation, just because the question of how to update *irrational* beliefs seems (and has usually been treated as) a separate question to rational update).

[9]It's interesting to contrast this to *pragmatic* arguments for conditionalization, which presumably will be able to exploit any lapses in conditionalization, whatever change in stance we might point to explain the change.

[10]This criticism of the Greaves-Wallace argument as an argument for a diachronic norm is made in (Easwaran, 2013), who credits Mike Titelbaum for the point.

Milne, Peter. 2008. 'Betting on fuzzy and many-valued propositions'. *In: The logica yearbook.*

Skyrms, Brian. 2006. 'Diachronic coherence and radical probabilism'. *Philosophy of science*, **73**, 959–96.

van Fraassen, Bas. 1999. Conditionalization, a new argument for. *Topoi*, **18**(2), 93–96.

Williams, J. Robert G. 2012a. 'Generalized probabilism: dutch books and accuracy domination'. *Journal of philosophical logic*, **41**(5).

Williams, J. Robert G. 2012b. 'Gradational accuracy and non-classical semantics'. *Review of symbolic logic*.

Williams, J. Robert G. forthcoming. 'Nonclassical logic and probability'. *In:* Hajek, Alan, & Hitchcock, Christopher (eds), *A companion to the philosophy of probability*.