Dr. Truthlove, or How I Learned to Stop Worrying and Love Bayesian Probabilities

Kenny Easwaran

10/15/2014

1 Setup

1.1 The Preface Paradox

Dr. Truthlove loves believing things that are true, and hates believing things that are false. She has just written an extensively researched book, and she believes every claim in the body of the book. However, she is also aware of the history of other books on the same subject, and knows that every single one of them has turned out to contain some false claims, despite the best efforts of their authors. Thus, one of the claims she makes, in the preface of the book, is to the effect that the body of this book too, like all the others, surely contains at least one false claim. She believes that too.

She notices a problem. At least one of her beliefs is false. Either some claim from the body of the book (all of which she believes) is false, or else the claim from the preface (which she also believes) is. So she knows that she's doing something that she hates — believing a false claim.

At the same time, she notices a benefit. At least one of her beliefs is true! Either the claim from the preface is true, or *all* of the claims in the body of the book are true. So she is doing something that she loves — believing a true claim.

But none of this answers the overall question. Is she doing what she ought to do? There is something apparently uneasy about her situation, but she can't be sure whether it's good or bad.

1.2 The Bayesian response

Some of Dr. Truthlove's colleagues notice her situation and propose the following sort of Bayesian account. They say that she is old-fashioned for thinking that there is an attitude of "belief" that plays an important role in epistemology.

Rather, they say, the appropriate way to think of her doxastic state is in terms of attitudes that come in degrees from 0 to 1. These "credences" are the things that matter.

Because these "credences" aren't all or nothing, because they come in degrees, she is being too hard on herself to evaluate them in terms of truth and falsity. They say that if one has a credence of .9 in one proposition and a credence of .7 in another, one wasn't "right" or "wrong" if either, both, or neither of these propositions turns out to be false. Instead, the only requirements that these "credences" have is that they must satisfy the axioms of probability, and they should be updated over time in accord with the rule of conditionalization on her evidence.[1]

However, they say that we don't have perfect introspective access to these credences, as evidenced by the fact that we still seem to talk in these old-fashioned terms of "belief" and lack thereof. In particular, when we have credences that are very high (say, above .99), we are likely to mistake this mental state for one of "belief". (Some of them instead say that having such a high credence just *is* what we mean by "belief", a view known as the "Lockean thesis". (Foley, 1993)) However, because of the way the probability axioms work, these degrees aren't preserved under conjunction. If one has credence .99 in each of two claims, this doesn't guarantee anything more than credence .98 in the conjunction. For three claims, one might have credence as low as .97. And with 100 claims, any credence in the conjunction is compatible with each conjunct having credence .99. Thus, as long as the body of her book has at least 100 claims, Dr. Truthlove might have credence above .99 in every claim in the body of the book, and yet also have credence above .99 that at least one of these claims is false, without violating the axioms of probability. If the threshold is higher than .99, then the number of claims required to reach this situation is correspondingly larger, while if the threshold is lower then then number required is smaller.

Thus, these Bayesians say, there is nothing puzzling about Dr. Truthlove's state at all. Once she gives up on the old-fashioned notion of belief, and a corresponding requirement of truth and falsity, she can accept the Bayesian solution and see that there is nothing wrong with her mental state at all, except for how she describes it.

1.3 Problems for Bayesianism

However, Dr. Truthlove doesn't buy it. This Bayesian "solution" seems to raise more questions than it answers.

First — what are these "credences"? Some Bayesians say that they just are one's dispositions to place bets. To say that one has credence x in a proposition is to say that one is disposed to be willing to pay any amount less than x for a bet that returns 1 unit of utility if the proposition is true, and to be willing to accept any amount greater than x for a bet that loses 1 unit of utility if the proposition is false. (Ramsey, 1926, de Finetti, 1974) Others recognize that we

[1] See, for instance, (Easwaran, 2011a) for discussion of what this entails.

don't always have these dispositions, and instead identify credences with some normative commitment to find these bets fair (Christensen, 1991) or with a sort of best systematization of our overall decision-making behavior. (Savage, 1954) However, some say that although credal states can help guide our actions, they must fundamentally be a different kind of thing. (Eriksson and Hájek, 2007)

Second — why should they obey the axioms of probability? Proponents of the betting interpretation give "Dutch book" arguments, and proponents of more general decision-theoretic roles for credence give "representation theorem" arguments. But for those that take a more purely epistemic standpoint on what credences are (Joyce, 1999, Jaynes, 2003) the arguments are more problematic, assuming either a notion of "closeness to truth" or uniqueness of what an evidential situation requires.

Third — the orthodox Bayesian theory says that credences are infinitely precise real numbers (Elga, 2010), but at least some Bayesians think that this level of precision in the human mind is implausible, and instead say that credences should be something less precise. (White, 2009) Of course, if they are less precise, then there are further questions about which of the many possibilities for mathematical representation is correct. (Walley, 1991)

Fourth — if there is some threshold above which credences act like "belief", then what is that threshold, and why does it take the value that it does?

Fifth — there is some level of idealization implicit in all of the relevant mathematical theories. Actual agents are likely to fall short of the ideal in a variety of ways. (Hacking, 1967) If people don't actually satisfy these axioms, then their credences may behave very differently from probabilities. In that case, what use is the Bayesian picture for describing actual agents?

There are of course further worries, (Easwaran, 2011b) but I will focus on these five here. The main thesis of this paper is that by accepting the old-fashioned notion of belief, and the simple view that the fundamental values of belief are to believe things that are true and not to believe things that are false, Dr. Truthlove can come to support something that sounds very much like the Bayesian solution to the Preface Paradox, while side-stepping all of these worries. "Credences" can be explained entirely in terms of the overall pattern of an agent's beliefs, and all the mathematical idealizations of the probability axioms, infinite precision, and thresholds can be accepted as merely a tool for summarizing the agent's beliefs, and her values for truth and falsity, which are not themselves committed to these further flights of mathematical fancy.

2 Formal background

To argue for this claim, I will need to set up the Truthlove situation in a more mathematically precise way. There are probably alternative ways to set things up that will allow the same conclusion (and in fact I will show two such ways to generalize things in Appendices B and C), but for now I will use a particular precisification. It is clear that many of the assumptions that go into this are somewhat implausible, but I hope that in future work some of these assumptions

can be weakened.

The first set of assumptions concerns the nature of the objects of belief. I will call the objects of belief "propositions", and I will assume that these propositions are characterized by sets of possible "situations". These situations are not metaphysically possible worlds — for instance, there may be situations in which Hesperus is not Phosphorus. Rather, these situations represent the uncertainties for the agent. A proposition that contains all the situations is one that the agent is certain of. For any proposition she is not completely certain of (whether she counts as believing it or not) there will be situations not contained in it. The set of situations is relative to the agent and her certainties; as a result, the same proposition may be characterized by one set of situations for me and a different set of situations for you.

The claims are that each proposition corresponds to a set of situations for each agent, that every set of situations for an agent corresponds to at least one proposition (and in fact, probably infinitely many), and that if two propositions correspond to the same set of situations for an agent, then the agent has the same attitude to them. (This last claim may be better interpreted as the contrapositive, stating that if an agent has distinct attitudes to two propositions, then there is at least one situation for the agent that is in one but not the other of the sets corresponding to these propositions.) Thus, although propositions themselves may well have much finer structure than these sets of situations, since these sets of situations will do all the work needed for me, I will use the word "proposition" to refer to a set of situations.

I will make some further assumptions about the nature of the doxastic state of an agent. First, I assume that the set of situations for each agent is finite. One might try to justify this assumption by arguing that actual agents are finite beings that only have the capacity or desire to make finitely many distinctions among ways the world could be. I happen to think that this sort of argument won't work, and that in fact the set of situations for a given agent is ordinarily infinite, but for the purposes of this paper I will restrict things to the finite case. Appendix A argues that my main results may not need to be changed too much.

Additionally, I will assume that the agent's doxastic state is fully characterized by the set of situations that are possible for her, together with the set of propositions that she believes. Some have argued that this is not enough, and we need disbelief as a separate attitude in order to characterize the concept of negation. (Wedgwood (2002) cites Rumfitt (2000) for this view.) Others have argued that in fact we also need an independent attitude of suspension of judgment beyond the mere lack of belief or disbelief. (Friedman, 2012) In Appendix C I will show that adding these notions as well gives the same results as just using belief, so that this assumption of mine is not an important one. At any rate, I think that both of these attitudes are only needed if one holds a more structured notion of proposition than a mere set of situations.

I won't make any further metaphysical assumptions about the nature of belief. I just assume that there is some meaningful state of mind that can be characterized as believing a proposition or not. In particular, I don't assume

anything limiting the possibility of any combination of beliefs — it is possible for an agent to believe both a proposition and its negation, or to believe a proposition and fail to believe something entailed by it, even on this characterization of propositions as sets of situations. (Note that on this characterization, the conjunction of two propositions is just the intersection of the sets of situations — thus, I am saying that it is perfectly possible to believe two propositions without believing their conjunction, or even to believe a conjunction without believing either conjunct.) However, it is important for Dr. Truthlove that the notion of belief does have some sort of reality, and that it is more fundamental than credence, if credence even is a real state at all.

I will make some very substantive normative assumptions about the evaluation of belief. In particular, I will spell out more precisely the "Truthlove" idea that the agent values believing things that are true and disvalues believing things that are false. The idea is that, no matter how things are, the agent is certain that exactly one of the situations is the actual one. An agent's doxastic state receives a value by counting up how many of her beliefs are true in this situation, and how many of her beliefs are false in this situation. I assume there is a value R that the agent gets for each belief that is right, and a value W that the agent loses for each belief that is wrong, and that the overall value of an epistemic state is the sum of the values for all the beliefs it contains. (This assumption is discussed further in Appendix D.) Importantly, propositions that the agent doesn't believe make no contribution to the overall value of a doxastic state, except insofar as the agent could have done better or worse by believing them. The value for being right or wrong is independent of the content of the proposition. (This assumption is weakened in Appendix G.)

Note that this view allows no room for considerations like evidence, consistency, coherence, justification, or anything else in the evaluation of a doxastic state, except insofar as such considerations can be explained in terms of truth. Along the way, I will show that such considerations are (at least to some extent) explained in terms of truth, as suggested by Wedgwood (2002). But where Wedgwood includes credence as well as belief as independent parts of a doxastic state, I will take belief as the only fundamental doxastic attitude, and argue that this version of the truth norm in fact allows credence to be reduced to belief.[2,3]

This picture is also inspired by the one proposed by William James in "The Will to Believe". (1896) In that lecture, James argues against Clifford ("The Ethics of Belief") who seeks a distinctive role for evidence. Although James is

[2]Wedgwood cites (Joyce, 1999) for a version of the argument involving credence. The present project was in fact inspired by trying to generalize Joyce's argument to the case of belief. Another version of this argument, which is subject to many of the same worries, is provided by (Leitgeb and Pettigrew, 2010a,b).

[3]The relation between a fundamental truth norm of this sort and these other norms of justification and evidence and the like is the subject of a dilemma presented on p. 150 of (Percival, 2002): either truth is not the sole fundamental value (in which case this whole argument collapses), or these other values are mere fictions that are useful for getting at truth (in which case we have to deny the value of these norms whenever they conflict with truth). The argument of this paper comes down on the second horn of this dilemma.

motivated by the thought that there are some propositions for which we can never have evidence unless we already believe them (in particular, he suggests examples of religious or social propositions, as well as cases of scientific propositions that we won't be motivated to investigate unless we already believe them) the basic picture is still similar. James suggests that the values I have called "R" and "W", representing the agent's value for truth and falsity of belief, are just fundamental features of an agent's "passional nature", that help determine whether she is highly committed or highly agnostic. We will see some of this in my formalism as well.

The substantive commitments of my picture that go beyond the proposals of James or Wedgwood are the claim that there is a mathematically precise value assigned to each belief, that this value is constant for each proposition, and that the overall value of a doxastic state is given by the sum of the values of the beliefs involved in it. There are clearly worries about each of these commitments, but rather than getting sidetracked by these worries (which are discussed to some extent in the appendices, but call for substantial further investigation), I will investigate the consequences of this picture, so that we can have a clearer idea of its costs and benefit before making a final judgment on it.

2.1 Examples

Most of the examples I discuss will concern doxastic states where there are just three possible situations. It is of course not plausible that many agents will ever be in such a state, but I use it because this state is the simplest one that allows me to illustrate many of the important features of the overall framework. Most of what I say generalizes to more complex states.

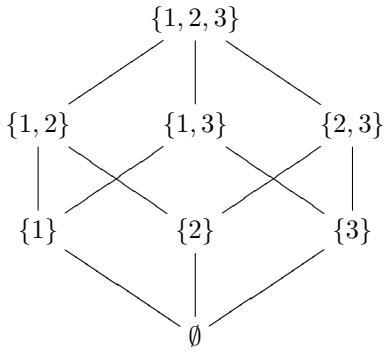


Figure 1: This diagram indicates the propositions that each position stands for in later diagrams.

I will illustrate doxastic states by means of diagrams like those in Figure 2. The way to read these diagrams is indicated in Figure 1. The three situations are named "1", "2", and "3". Each proposition corresponds to a set of situations, and the lines indicate entailment relations between propositions. The place at the top of the diagram represents the proposition that is true in all three situations, the ones on the second line represent the propositions true in exactly two situations, the ones on the third line represent the propositions true in exactly one situation, and the one on the bottom represents the impossible proposition. If a position contains "B", then it indicates that the doxastic state involves belief in that proposition. "$*$" is a placeholder for propositions that are not believed.

The doxastic state in Figure 2 has several strange features. Note that the agent believes both the proposition true only in situations 1 and 3, and also

the proposition true only in situation 2, which are negations of each other. For another pair of a proposition and its negation, $\{2,3\}$ and $\{1\}$, this agent believes neither. The overall summary of the agent's doxastic state needs to say of each proposition individually whether the agent believes it or not — it is not sufficient to say what the agent believes about the situations themselves.

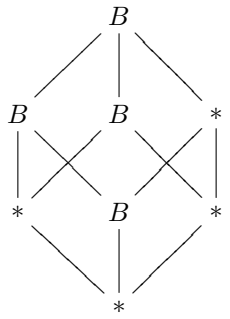


Figure 2: An example doxastic state. Score: $(3R-W, 3R-W, 2R-2W)$

To evaluate a given doxastic state, we have to see which beliefs in it are true and which are false. But this of course depends on which situation is actual. We can read this off the diagram by following the lines. For instance, if situation 1 is actual, then the proposition in the lower left of the diagram is true, as well as everything that can be reached from it by following lines upwards. All other propositions are false in situation 1. Similarly, if situation 2 is actual, then the proposition in the lower middle is true, as well as everything that can be reached from it by following lines upwards, and all other propositions are false. Similarly for situation 3 and the proposition in the lower right.

Thus, for the doxastic state illustrated in figure 2, in situation 1 the state has 3 beliefs that are right and 1 that is wrong, in situation 2 it again gets 3 right and 1 wrong, and in situation 3 it gets 2 right and 2 wrong. This is what I mean when I say that its "score" is $(3R-W, 3R-W, 2R-2W)$. Of course, the actual numerical values depend on the numerical values of the parameters R and W.

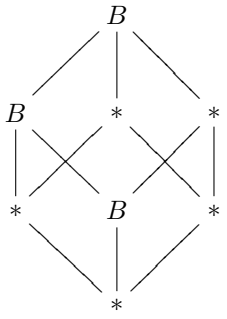


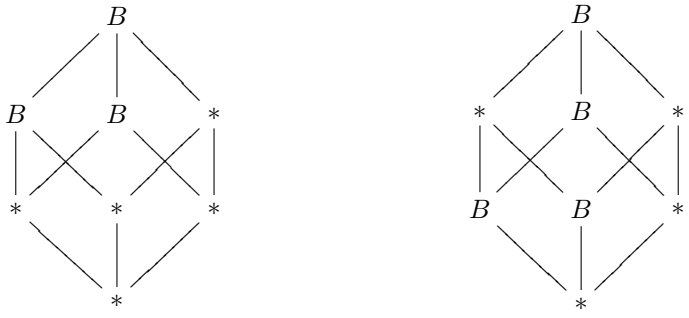
Figure 3: Score: $(2R-W, 3R, R-2W)$

If we compare this to the doxastic state in Figure 3, we can see that the one from Figure 2 is better in situations 1 and 3, while the one from Figure 3 is better in situation 2. Thus, to figure out which doxastic state is better, one needs to know which situation is actual. Since the agent doesn't know which situation is actual, the agent can't figure this out herself.

2.2 Dominance and Coherence

However, consider what happens with the doxastic states shown in Figure 4a and 4b. For these two doxastic states, it doesn't matter which situation is actual. The one in 4a *always* gets a better score than the one in 4b.

Borrowing terminology from decision theory, I will say that 4a *dominates* 4b. In general I will say that one doxastic state *strongly dominates* another iff they are defined over the same set of situations, and the former has a strictly higher score in every single situation. I will say that one doxastic state *weakly dominates* another iff they are defined over the same set of situations, and the former has at least as high a score in every single situation, and there is some



(a) $(3R, 2R - W, 2R - W)$ (b) $(3R - W, 2R - 2W, 2R - 2W)$

Figure 4: The left one dominates the right one.

situation in which the former has a strictly higher score.

The idea is that in the case of strong dominance, one can already be sure that the dominating doxastic state is better than the dominated one, and in the case of weak dominance, one can be sure that the dominated one is no better than the dominating one, and the latter may in fact be better depending on which situation is actual.

To go along with these notions, I will define notions of "coherence". I will say that a doxastic state is *strongly coherent* iff there is no doxastic state that weakly dominates it. I will say that a doxastic state is *weakly coherent* iff there is no doxastic state that strongly dominates it. Any strongly coherent doxastic state is also weakly coherent, but the converse is not true. Given the normative role the scores play, it seems that we can say the following:

> **Strong Coherence:** A rational agent must have a doxastic state that is strongly coherent.[4]

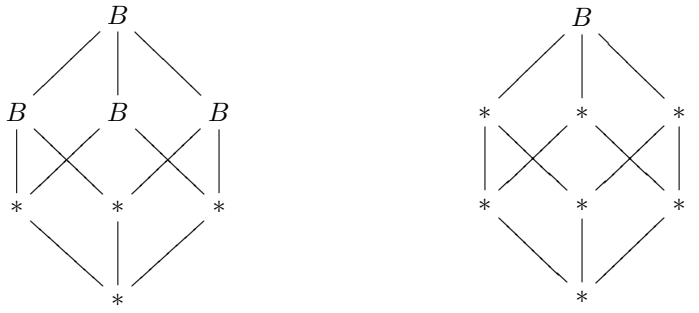
If she doesn't, then there is another doxastic state that she can recognize as being an alternative that will always do at least as well, and may in fact do better.

This rule doesn't say anything about how to repair one's beliefs if they are not strongly coherent. It doesn't say that one must switch to one of the doxastic states that dominates one's current doxastic state. It just says that if some other doxastic state weakly dominates the one that you currently have, then you are doing something wrong. Perhaps you ought to instead have one of the ones that dominates your current state. Or perhaps you ought to have some completely different doxastic state, apart from the one that dominates your current state, though the one that you ought to have clearly should not be dominated by any other state.

This rule also doesn't say anything about *which* coherent doxastic state the agent should have. I will show later that there are many coherent doxastic

[4]It is possible to formulate a version requiring only weak coherence, but it seems to me that such a view would be under-motivated. In Appendix E I show (among other things) that states that are weakly coherent without being strongly coherent are quite strange, and are not plausibly rational.

states, and for all that I say, there may be some that run afoul of some other condition. The requirements of rationality may involve more than just **Strong Coherence**. In the next section I will discuss some necessary conditions for being coherent, and some sufficient conditions for being coherent, to show more about what this requirement amounts to, and eventually I will come back to the question of whether rationality might require more as well.



(a) $(3R - W, 3R - W, 3R - W)$ (b) (R, R, R)

Figure 5: Which one dominates depends on R and W.

But I will make one final point. In all the previous examples, whether or not dominance occurred didn't depend on the values of R and W. But there are some cases in which it does. For instance, in Figure 5, the doxastic state on the left always gets score $3R - W$, while the one on the right always gets score R. Thus, if $W > 2R$, then the one on the right dominates, while if $W < 2R$ then the one on the left does.

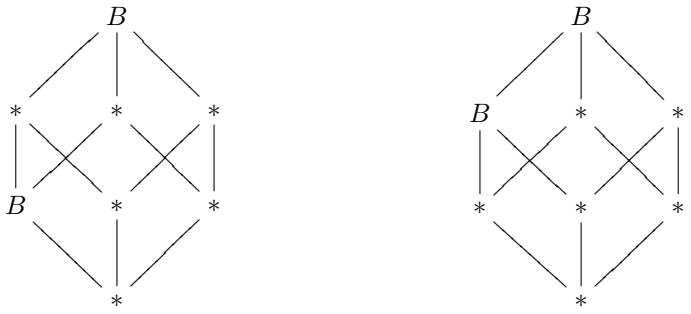
As it happens, whichever one of these dominates the other, it will turn out to be strongly coherent. But to show this, I will need to examine the conditions for dominance and coherence more closely.

3 Conditions for coherence

3.1 Necessary conditions

The first conditions I will discuss are some logical conditions. Consider what happens if there is a proposition p such that the agent believes both it and its complement (which I will call "$\neg p$"). In every possible situation, exactly one of these will be wrong and the other will be right. Thus, these two together always contribute $R - W$ to the overall score. If the agent had instead not believed either, then these two propositions would always contribute 0 to the overall score. Thus, if $R < W$, we can see that in order to be (strongly or weakly) coherent, a doxastic state must not include belief in both p and $\neg p$. Oddly, if $R > W$, then in order to be (strongly or weakly) coherent, a doxastic state must always include belief in at least one of p or $\neg p$. This consequence is strange enough that I will stipulate that for any agent, $R \leq W$. (In Appendix B, I discuss a generalization of the overall framework that eventually shows that

this is not actually a substantive assumption, but rather a terminological one, if we assume that attitudes are individuated by their normative functional roles.)



(a) $(2R, R-W, R-W)$ (b) $(2R, 2R, R-W)$

Figure 6: p is $\{1\}$, q is $\{1,2\}$. (b) weakly dominates (a), because of the score in situation 2.

Another logical requirement on belief involves entailment, and is illustrated in Figure 6. Imagine that p and q are two propositions, with p a proper subset of q when interpreted as sets of situations, so that p entails q (either logically, or from the point of view of the agent) and is not equivalent to it. Consider an agent who believes p but not q, and a different agent who has all the same beliefs, except that she believes q but not p. For any situation that is a member of p (and thus q as well), the two agents have the same overall score, because they get the same score from all other propositions, and the belief the agent has in one of p and q is correct. For any situation that is not a member of q (and thus not of p either), the two agents also have the same overall score, because this one belief is false, and the other propositions get the same score. So any situation in which the two agents have a different score is a member of q that is not in p. And in this case, the second agent has a strictly better score than the first.

Thus, if p entails q, and one believes p but not q, then one's doxastic state is weakly dominated, and thus not strongly coherent. Therefore, in order to be strongly coherent, a doxastic state must be such that it believes all the consequences of any single proposition that it believes. We get a sort of single-premise closure requirement for belief.

Importantly, **Strong Coherence** only supports "wide scope" versions of these rules. That is, it supports: one ought to be such that (if one believes p then one believes q) — it does not support: if one believes p then (one ought to believe q). The latter would require some further claim about *which* strongly coherent doxastic state one ought to have.

Niko Kolodny suggests, in his (2007), that fundamentally, there are no such wide scope norms on belief. Instead, he argues that all such apparent norms are the result of various narrow scope norms on individual beliefs. For instance, he suggests that in the case of the apparent wide scope norm not to both believe p and believe $\neg p$, what is really going on is that by believing both, one is violating the norm to only believe propositions that are supported by one's evidence —

at most one of these two propositions can be supported by one's evidence. On the present proposal, we can say something similar, though not identical. The norm not to both believe p and believe $\neg p$, and the norm not to believe p while failing to believe something entailed by p, both arise from an overall evaluation of doxastic states in terms of the value of each individual belief being true or false. This purely local evaluation of doxastic states gives rise to a global norm relating separate beliefs.

At this point we have seen that there are some logical requirements that a doxastic state must satisfy in order to be strongly coherent. (It is straightforward to also show that in order to be weakly or strongly coherent, it is necessary to believe the proposition that is true in every situation, and not to believe the proposition that is true in no situations.) I will now turn to some sufficient conditions for coherence, which will in turn let me show that some other plausible logical requirements aren't actually necessary.

3.2 Sufficient conditions

The first way to guarantee coherence is to choose some situation and believe all and only the propositions that are true in that situation. The version for situation 2 is illustrated in Figure 7. Having such a doxastic state will guarantee that one gets the highest possible score if the relevant situation is actual. (If there are n situations, then there are 2^n propositions, and half of them are true in any given situation, so the overall score in that situation is $2^{n-1}R$.) No distinct doxastic state can get a score this high in this situation, and thus no doxastic state can (either weakly or strongly) dominate it, and thus such a state is strongly coherent. (It is not hard to calculate that this doxastic state always gets score $2^{n-2}(R-W)$ in any situation other than the one where it is maximally correct — it will always be the case that half of these beliefs are right and half wrong in any other situation.)

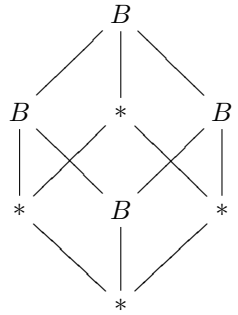


Figure 7: $(2R - 2W, 4R, 2R - 2W)$

But as long as there are at least two possible situations, these will not be the only strongly coherent doxastic states. To demonstrate more of them, I will first introduce some further tools from decision theory.

Dominance arises as a criterion for decision making in situations of uncertainty. If one has multiple options, and the values of those options depends on some feature of the world that one doesn't know, then one may not know which option is best. However, if one option is better than another in every situation, then the second is clearly not the option to take. (Again, the first may not be the right one either, depending on what other options there are, but we can use dominance to eliminate the second.) But dominance is not the only rule for decision making under uncertainty. In particular, most decision theories consider "expected value" important in making such decisions. I won't actually assume any normative force for expected value itself, but I will use it

as a tool for clarifying when dominance does or doesn't occur, and thus use it to show that certain belief sets are (weakly or strongly) coherent.

Expected value is defined in terms of a probability function P. To say that P is a probability function is just to say that it assigns a non-negative real number to each proposition, such that $P(p \cup q) = P(p) + P(q)$ whenever p and q are disjoint (incompatible), and such that $P(p) = 1$ if p is the proposition that includes all possible situations. If s is a situation, then we often abuse notation and write "$P(s)$" to mean $P(\{s\})$, where $\{s\}$ is the proposition that is true only in s. Because of these stipulations (and the fact that there are only finitely many situations), we can see that a probability function P is determined entirely by the values $P(s)$, which must themselves be non-negative real numbers summing to 1.

If f is a function that is defined on situations, then we define the expected value of f with respect to P to be

$$E_P(f) = \sum f(s) \cdot P(s),$$

where the sum ranges over all situations. As a result, we can see that if $f(s) < g(s)$ for all situations s, then $E_P(f) < E_P(g)$, for any probability function P. (This is because $f(s) \cdot P(s) \leq g(s) \cdot P(s)$, with equality occurring only if $P(s) = 0$. Since at least some situation must have $P(s) > 0$, at least one term in the sum for g is strictly greater than the corresponding term in the sum for f.)

Since the score of a doxastic state is itself a function that is defined on situations, we thus get the following theorems:

> If doxastic state A strongly dominates B, then for any probability function P, the expected score of A on P is strictly greater than the expected score of B on P.
>
> If doxastic state A weakly dominates B, then for any probability function P such that $P(s) > 0$ for all s, the expected score of A on P is strictly greater than the expected score of B on P.

The theorems are proved by first noting that every term in the sum for the expected score of A is at least as great as the corresponding term in the expected score for B. The only way for the two terms to be equal is if s is a situation where either the score of A is exactly equal to the score of B (which is never possible if A strongly dominates B) or where $P(s) = 0$ (which is never possible under the conditions of the second theorem). Thus, under either set of conditions, at least one of the terms for A is strictly greater than the corresponding term for B, so the expected score of A is strictly greater than that of B.

We will actually use the converse of these theorems. If we can find a probability function P such that A has maximal expected score for P, then the first theorem tells us that no other doxastic state strongly dominates A, and if P doesn't have $P(s) = 0$ for any situation, then the second theorem tells us that no other doxastic state even weakly dominates A. Thus, by choosing appropriate

P and finding doxastic states that maximize expected score for P, we can find weakly coherent doxastic states, and they will in fact be strongly coherent as long as every situation has non-zero probability. So this just leaves the project of finding a doxastic state that maximizes expected score for a given probability function.

The feature of expected value that will help here is what is known to mathematicians as "linearity of expectations". That is,

$$E_P(f+g) = E_P(f) + E_P(g).$$

To show this, we note that

$$E_P(f+g) = \sum((f(s)+g(s)) \cdot P(s)).$$

But by rearranging the terms in the sum, we see that

$$\sum((f(s)+g(s)) \cdot P(s)) = \sum(f(s) \cdot P(s)) + \sum(g(s) \cdot P(s)) = E_P(f) + E_P(g).$$

Since the score of a doxastic state in a given situation is the sum of the scores of the individual beliefs that make it up in that situation, we can use the linearity of expectations to conclude that the expected score of a doxastic state is the sum of the expected scores of the individual beliefs that make it up.

Thus, for a given probability function, we can find a doxastic state that maximizes expected score just by figuring out, for each proposition, whether believing it or not believing it has a higher expected score, and then choosing a doxastic state that has the maximizing attitude (for this probability function) to each proposition. So the question of finding a doxastic state that maximizes expected score for a probability function comes down to figuring out which attitude to a given proposition maximizes expected score for a given probability function.

And in general, it is not hard to show that the expected score of believing p will be $R \cdot P(p) - W \cdot P(\neg p)$ (since $P(p)$ is the sum of $P(s)$ for all s in p, and the score for each such s is R, and similarly for $\neg p$ and $-W$), and the expected score of not believing p will be 0 (since that is the score of non-belief in *every* situation). Thus, for a given probability function, to maximize expected score, a doxastic state must believe p if $R \cdot P(p) > W \cdot P(\neg p)$, and must not believe p if $R \cdot P(p) < W \cdot P(\neg p)$. If $R \cdot P(p) = W \cdot P(\neg p)$, then both attitudes are equally maximal in expected score. After noting that $P(\neg p) = 1 - P(p)$, we can see that the condition comes down to the following:

> For a given probability function P, a doxastic state maximizes expected score iff it believes all propositions p such that $P(p) > \frac{W}{R+W}$ and believes no propositions p such that $P(p) < \frac{W}{R+W}$. Both believing and not believing are compatible with maximizing expected score if $P(p) = \frac{W}{R+W}$.

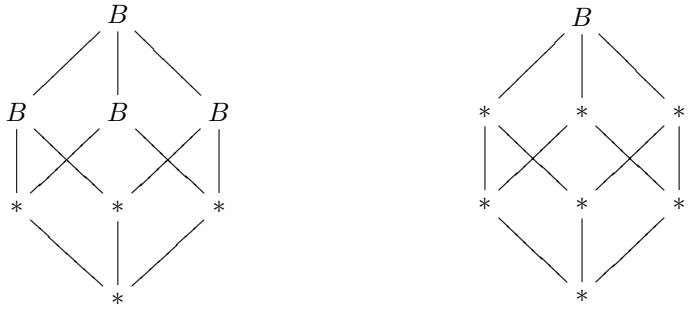
If a doxastic state A bears this relation to a probability function P, then I will say that P *represents* A.

Thus, putting this together with the earlier results, we can draw the following conclusions:

> **First Main Result:** If A is a doxastic state that is represented by a probability function P, then A is weakly coherent.
>
> **Second Main Result:** If A is a doxastic state that is represented by a probability function P, and if $P(s) > 0$ for every situation s, then A is strongly coherent.

This lets us generate many strongly coherent doxastic states by considering various probability functions, and seeing how the values relate to the threshold $W/(R+W)$.

As an example, consider the "uniform" probability function on three situations, such that $P(1) = P(2) = P(3) = 1/3$. The empty proposition has probability 0, the three one-situation propositions have probability $1/3$, the three two-situation propositions have probability $2/3$, and the full proposition has probability 1. Since we have assumed that $W \geq R$, which doxastic state is represented by this probability function just depends on how $W/(W+R)$ compares to $2/3$, since it will definitely be between $1/2$ and 1. This is illustrated in Figure 8. (Note that these are exactly the doxastic states illustrated in Figure 5 above.)



(a) $(3R-W, 3R-W, 3R-W)$ (b) (R, R, R)

Figure 8: If $W \leq 2R$ then the left is represented by the uniform probability function, and if $W \geq 2R$ then the right is.

One thing to note about the doxastic state illustrated in Figure 8a is that such an agent has inconsistent beliefs. It is not possible for all four propositions that she believes to be true, and yet her set of beliefs is strongly coherent — there is no other doxastic state that is guaranteed to be better in terms of getting beliefs right or wrong. We showed in section 3.1 that in order to be coherent, it is necessary that one not believe a *pair* of inconsistent propositions, and it is necessary that one believe the consequences of any *single* proposition that one believes. This example shows that (for certain values of R and W) it is not necessary to avoid believing an inconsistent *triple* of propositions, and it is not necessary to believe the consequences of a *pair* of propositions that one believes.

Note that, as William James suggested, the amount of commitment one finds in a coherent doxastic state may depend in general on the relative values the agent has for getting things right or wrong. If getting things wrong is extremely bad, then a very non-committal doxastic state can be coherent. But if getting things right is comparatively important, then this state will be incoherent, and will be dominated by some more committed doxastic states. Of course, no matter what the relative values are like, there will always be some maximally committed doxastic states that are coherent (namely, the ones like in Figure 7, where one believes all and only the propositions that are true in a particular situation). But, no matter what the values are, and no matter how many situations there are, a uniform probability function will always represent a strongly coherent doxastic state that is not one of these maximal states, and various non-uniform probability functions will often represent others as well.

4 Return to Truthlove

4.1 Improvements on the Bayesian account

This result allows Dr. Truthlove to adopt a version of the Bayesian solution to the Preface Paradox. The Bayesians say that in reality, Dr. Truthlove doesn't have "beliefs", but instead she has credences that satisfy the probability axioms, and the propositions she claims to "believe" are just the ones that have sufficiently high credence. By means of the **Second Main Result**, Dr. Truthlove can take the probability function the Bayesians attribute to her, and use it to show that her doxastic state is strongly coherent. If the Bayesians set the threshold credence to be t, then she can choose R and W so that $W/(R+W) = t$[5]. They say that the propositions whose probability are above the threshold are precisely the ones she believes, and thus the probability function represents her doxastic state, which is thus strongly coherent.

Interestingly, where the Bayesians say that the probability function is the real doxastic state, and the "beliefs" are just a verbal representation of it, Dr. Truthlove can say that the *beliefs* are the real doxastic state, and the probability function is just a numerical representation of it. The probability function isn't taken to be any real feature of the agent — it is just a device for showing that the beliefs are strongly coherent. Strong coherence is just the claim that no other doxastic state can do as well in every possible situation as this one, when evaluated just in terms of truth and falsity. Since this probability function is only a mathematical tool, and not an actual feature of the doxastic state, this

[5]She can always do this because

$$\frac{W}{R+W} = \frac{1}{R/W+1},$$

and thus by setting $R = W(1/t - 1)$, she can guarantee that

$$\frac{1}{R/W+1} = \frac{1}{(1/t-1)+1} = t.$$

approach avoids the first problem raised for Bayesianism in section 1.3, namely the problem of saying what credences are. And in fact, it avoids the other problems raised there as well.

For the question of why credences should obey the axioms of probability, there is no problem. The probability function is just a tool for showing that certain doxastic states are coherent. I used a probability function because it allowed me to adopt the tools of expected value and decision theory, but if some other sort of function had the same sort of connection to coherence, it would be acceptable as well.

For the question of whether the probability function should be infinitely precise, or should instead be interval-valued, or otherwise imprecise, the lack of realism again provides a response. If I claim to believe a proposition, then the orthodox Bayesian must say that in fact I actually have some particular credence that is greater than t. However, Dr. Truthlove can deny that any particular precise value is *the* probability of that proposition for the agent. As shown in Appendix E, if there is a probability function representing a doxastic state for which no proposition has probability *exactly* equal to $W/(R+W)$, then there are in fact infinitely many probability functions that equally well represent this same doxastic state. The only fact about the probability function that is relevant is how the probability of each proposition compares to $W/(R+W)$, and thus any of these probability functions that represents the doxastic state is equally good. In the cases I have been considering, where there are only three situations, the range of probability values that are compatible with representation are often quite wide. But if there are far more situations, and the agent has a very complicated pattern of beliefs, then the set of representing probability functions may take on much narrower ranges of values.

Given this range of functions, one might deny that any of the functions is itself an appropriate representation of the agent's doxastic state. Rather, one might say that she is represented by the set of all these probability functions, or the intervals of values these probability functions take on. But there is no real need to decide whether probabilities are *really* precise, or that they are *really* imprecise. *Really*, there are just beliefs, and the probability function is only a mathematical tool for summarizing certain facts about the beliefs.

And in fact, this is exactly what Bayesians who use a decision-theoretic argument already should say. Jeffrey (1965) is explicit that there is no unique probability and utility function that represents an agent, and Dreier (1996) proposes that this means that utilities and probabilities should be taken in precisely this anti-realist way proposed here. Zynda (2000) and Meacham and Weisberg (2011) try to use this multiplicity of representation as a criticism of the representational approach. But I think it is better to use this multiplicity to focus attention on the beliefs (in the present case) or preferences (in the case of decision theory) and to think of the probabilities (and utilities, in the case of decision theory) as mere mathematical conveniences. They may have some sort of reality, but they are constituted by the more fundamental on-off states (belief or preference) rather than as fully fine-grained numerical things with their own independent mental existence.

Fundamentally, we know that the human mind is constituted by a mass of neurons and their interconnections with one another, and perhaps various other features of the body and the external world. If humans were actually robots, with minds that were constituted by inscriptions of sentences with corresponding probability numbers, then there would be reason to take the probability function more seriously. But at best, the probability function is some sort of emergent description of massively complex behavior. (Of course, the same is true for the notion of belief itself, which Dr. Truthlove takes as a primitive — there are very deep and interesting questions to consider here about what the nature of belief itself actually is, but these are far beyond the scope of this paper. If belief itself is just a rough approximation of what real doxastic states are like, then the foundations of this paper may need to be revised.)

Bayesians sometimes object that there are certain propositions that we do have infinitely precise credences in. For instance, if I am about to flip a coin that I believe to be fair, then it seems plausible that my credence that it will come up heads is *precisely* 1/2. However, Dr. Truthlove can deny this. She *believes* that the *chance* that the coin will land heads is precisely 1/2. But for the proposition that the coin will land heads itself, she merely fails to believe it. There are other propositions that are logically related to this one, some of which she believes and some of which she doesn't, and the pattern of her beliefs on these propositions provides some constraint on what a probabilistic representation of her doxastic state is like. But there is no precise number that is *the* credence in the proposition. She says that every case where Bayesians can make a case that there really is a precise degree of belief is actually a case like this, where there is a belief about a precise objective probability of some sort, which is being mistaken for a degree of belief.

Some Bayesians might say that even if we lack precise numerical degrees of belief, we at least have some notion of comparative confidence. I might be more confident that the laws of physics are the same throughout the universe than I am that life exists elsewhere in the universe. The Bayesian can explain this in terms of different numerical degrees of belief. However, Dr. Truthlove can also explain this by pointing to features of my overall doxastic state. An agent who believes both p and q may also believe $p \cap r$ but not $q \cap r$. An agent who believes neither p nor q may believe $p \cup r$ but not $q \cup r$. According to Truthlove, this is exactly how greater confidence in p than q will be manifested. And for agents with coherent doxastic states of this sort, probability functions that represent their state will generally assign higher values to p than to q. (Further research is needed to figure out precisely what conditions on belief lead to requirements that *every* representing probability function assigns a higher value to p than to q.) But of course, there will be many propositions that can't be compared in this way on the basis of an agent's doxastic state, and this seems right — it seems quite plausible that we have some pairs of beliefs that don't come with a useful comparison of confidence.

The fourth worry for Bayesianism that was raised in section 1.3 was the question of why the particular threshold involved is significant. On Dr. Truthlove's picture, the threshold is immediately derived from the agent's own values as

$W/(W+R)$. Of course, there is a further worry about what makes it the case that W and R themselves properly represent the norms on doxastic states, and whether and to what extent different agents really can have different values. (See Appendix D.) But this seems more like the sort of question one can get a grip on than the question of why some particular numerical threshold is important for the pre-theoretic concept of belief. The question is what might make it the case that one agent values not being wrong 5 times as much as being right while another agent values it only 1.5 times as much. James put this down to fundamental differences in our "passional natures", but surely more can be said.

Finally, the fifth worry for the Bayesian solution addressed the problem of non-idealization. We know that actual agents are not logically omniscient, and thus will fail to have credences that satisfy the probability axioms. A variety of modifications have been proposed (Hacking, 1967, Garber, 1983, Gaifman, 2004), which may have normative force, but all maintain too much omniscience to be adequate as descriptions of actual agents. On the Truthlove picture, this is not a problem. An ideal agent will always have a doxastic state that is strongly coherent, and may indirectly be associated with a probability function (or set of probability functions) that represents it. But a non-ideal agent may just have a set of beliefs that is not strongly coherent. Such a set of beliefs won't correspond in any natural way to a probability function, and there may be no natural way to associate numerical functions that fall short of the probability axioms either. But this is no problem — the beliefs themselves are what we care about, and the numerical representation is just a tool for calculation, when it exists. If we can say something further about what doxastic state a non-ideal agent ideally *should* have, then we may be able to say something further about probability functions associated with this ideal doxastic state.

5 Questions for further research

5.1 Non-representable coherent states

I have shown, with the two **Main Results** that if a doxastic state is represented by a probability function, then it is coherent. The claim that would be nice to have is the converse, that if a doxastic state is strongly coherent, then it is represented by a probability function. If this claim were true, then Dr. Truthlove would have a full justification of all applications of the Bayesian framework, and not just the preface paradox.

However, with some help from XX I have found that there are some doxastic states that are strongly coherent, but are not represented by a probability function. It turns out that the smallest such examples require four possible situations, and they require particular values of R and W. These examples are discussed in Appendix F.

Interestingly, when there are four situations, then if $W > 3R$, then for every strongly coherent doxastic state, there is a probability function that represents it. If a similar cutoff exists for larger numbers of situations, then we can get a

version of the converse. Just as I have required already that $W \geq R$, perhaps one can give an argument that $W \geq kR$, for some sufficiently large k, and in that case one can show that if a state is strongly coherent, then it is represented by a probability function.

Alternately, I have so far only assumed that a doxastic state ought to be strongly coherent. Perhaps there is some further norm on doxastic states that can be extracted from the Truthlove paradigm. If that is right, then this further norm may be sufficient to guarantee the existence of a probabilistic representation. (There may be ways to adapt the decision-theoretic representation theorems of Savage (1954), Jeffrey (1965), or Buchak (2013) for this purpose.) An example of a generalization of the norm that works for the case of four situations is given in Appendix G.

In the absence of these generalizations though, the best I can say is this. By finding a probability function and adopting the doxastic state that it represents, one can guarantee that one's doxastic state is strongly coherent. This may be the simplest and most general way to guarantee that one's doxastic state is strongly coherent, and so agents have some *prima facie* reason to have a doxastic state that is represented by a probability function, even though **Strong Coherence** only requires that they have *some* doxastic state that is strongly coherent. Thus, for now I get only a weak argument for the applicability of the Bayesian picture, but a proof of the conjecture from Appendix G would strengthen it.

5.2 Learning

Some further natural questions arise when considering diachronic features of a doxastic state. An important aspect of Bayesianism that hasn't been mentioned so far is the rule of update by conditionalization. This rule says that when an agent learns an evidence proposition e, she should change her credences so that her new credence in each proposition p is $P(p \wedge e)/P(e)$, where P is her old credence function. There is also a modification of this update procedure advocated by Richard Jeffrey, for cases in which no proposition is learned with certainty.

Gilbert Harman, among others, has argued that there can be no comparable rule for update of full beliefs. However, there are others (Alchourròn et al., 1985) that do try to give such a formal rule. A natural question to ask is whether the Truthlove paradigm can be used to justify such a rule, or to argue for a permissive alternative as Harman suggests. At any rate, Lin and Kelly (2012) provide interesting discussion on the prospects for such an update rule that is compatible with Bayesian conditionalization when there is a probability threshold for belief.

5.3 Change in value

Another interesting question, raised by Ian Hacking at an earlier presentation of this paper, is that of how beliefs should change in light of a change of values. If it is permissible for different agents to have different values of R and W,

then perhaps it can be permissible for these values to change for a given agent over time, even in the absence of new evidence. As we have seen earlier, which doxastic states are coherent depends on these values, and so after such a change in values, the agent may find herself with an incoherent doxastic state. Which doxastic state should she adopt as a result?

One natural suggestion would be to use the probabilistic representations. If there is a probabilistic representation of the agent's doxastic state before the change in value, then perhaps what she ought to do is adopt the doxastic state represented by this same probability function with the new values.

However, since the probabilistic representation is not unique, different probability functions that represent the same doxastic state before the change in values will often end up representing different doxastic states after the change. Perhaps this gives the outline of a permissive rule for doxastic change. Or perhaps it is just a problem for the whole approach. But it may also be part of a further argument for the existence of probabilistic representations — maybe the only systematic way to update doxastic states in light of change of values involves probability functions, and so we can require that agents have doxastic states that are able to make use of this method.

6 Conclusion

There are a variety of arguments that have been given for Bayesianism that presuppose the existence of some sort of doxastic attitude that comes in degrees. However, many philosophers are suspicious of such a presupposition, and support this suspicion with the fact that our ordinary concepts seem to be on-or-off rather than coming in degrees. But if the Truthlove idea of evaluating these full belief states is right, then there is often a way to find corresponding probability functions, and use them to simplify formalizations of our doxastic states. Thus, philosophers who are suspicious of the Bayesian program can still make use of it, without incurring any of the costs involved in assuming a real mental state that comes in degrees.

A Infinite Sets of Situations

I think the overall project of this paper can be extended to cases where the set of situations is infinite. However, some of the details need to be modified.

The basic elements of the main picture are the following. There is a finite set of situations, and the propositions are the subsets of this set. The agent believes some subset of these propositions. There are scores R and W that the agent gets for believing a proposition when it is true or false. The overall score of a doxastic state is the sum of the scores for the individual attitudes making up that state. A doxastic state is (strongly or weakly) coherent iff no other doxastic state (weakly or strongly) dominates it in overall score. If a doxastic state is strongly coherent, then if p entails q, it does not believe p without believing

q, and it does not believe both p and $\neg p$. A probability function represents a doxastic state iff every proposition with probability above $W/(R+W)$ is believed by that state, and none with probability less than $W/(R+W)$ is. If there is a probability function that represents a doxastic state, then the doxastic state is coherent.

One thing to note is that instead of considering the complete algebra of all propositions, we can restrict consideration to an *agenda*, which we can think of as a subset $\mathcal{A}$ of the set of all propositions. We can say that the score of a doxastic state on an agenda is the sum of the scores for the individual attitudes the state has to propositions in that agenda. We can say that a doxastic state is (strongly or weakly) coherent on an agenda iff no other doxastic state (weakly or strongly) dominates it in score on that agenda. We can say that a probability function represents a doxastic state on an agenda iff every proposition in that agenda with probability above $W/(R+W)$ is believed by that state, and none in that agenda with probability less than $W/(R+W)$ is.

Once we have these definitions, we can prove the following claims when the set of situations is finite. A doxastic state is coherent iff it is coherent on every finite agenda. (One direction is trivial, because the set of all propositions is a finite agenda, but the other direction is also true, since if state S' dominates S overall, then if we let $\mathcal{A}$ be the set of propositions on which the attitude of S' differs from that of S, then S' dominates S on $\mathcal{A}$.) A probability function represents a doxastic state iff it represents it on every finite agenda.

When the set of situations is infinite, we can no longer define the overall score of a doxastic state, since there may be infinitely many positive terms of size R and infinitely many negative terms of size W to sum. However, we can still define the score on a finite agenda in the same way as before, and we can then use the results of the preceding paragraphs as *definitions* of what it means for an overall doxastic state to be coherent or represented. We get the same arguments for single premise closure and avoidance of blatant contradictions. If a doxastic state S is strongly coherent on an agenda $\mathcal{A}$, then if p entails q, and p and q are in $\mathcal{A}$, then S does not believe p without believing q. If a doxastic state S is strongly coherent on an agenda $\mathcal{A}$, and if p and $\neg p$ are in $\mathcal{A}$, then S does not believe p and $\neg p$. And of course, the definition of what it takes for a probability function to represent a doxastic state does not need any modification based on the agenda being finite or infinite.

If further notions of coherence are developed that do entail probabilistic representability, then we will need to check if these notions of coherence can also be defined for infinite sets by means of defining them for finite agendas. Ideally, we will show that if a doxastic state satisfies this sort of coherence on a given finite agenda, then there is a probability function that represents it on that finite agenda. Given these claims, we can then show that if an overall doxastic state is coherent in this sense (so that it is coherent on every finite agenda, and thus represented on every finite agenda), then there is a probability function that represents it overall. (I owe this argument to XX.)

This is by an application of the Compactness Theorem from model theory. We can take the language of real numbers, add a constant t for $W/(R+W)$, and

a constant x_p for the probability of each proposition p. We then consider the theory including the full theory of real numbers, enough axioms to guarantee that t is above every rational below $W/(R+W)$ and below every rational above $W/(R+W)$, each instance of the probability axioms (such as: $x_p \leq 1$, $x_p \geq 0$, $x_p + x_q = x_{p \vee q}$ for each disjoint p and q, etc.), and sentences $x_p \geq t$ for propositions that are believed and $x_p \leq t$ for propositions that are not believed. Each finite subset of these sentences has a model by the existence of representations on finite agendas. And thus by the Compactness Theorem, the whole set of sentences has a model. This model may assign non-standard "hyperreal" values to some probabilities, but by taking the standard part of each probability value, we will get a standard real-valued probability function that represents the overall doxastic state. (We can't guarantee that the resulting function is regular or definite, in the sense of appendix E, since taking the standard part may make some strict inequalities weak.)

B The Universality of Belief

B.1 Introduction

In this appendix, I will show how to generalize my assumption that the agent's doxastic state can be described just by saying which propositions the agent believes and which she doesn't believe, provided that it is evaluated only in terms of truth and falsity of propositions. In fact, if a doxastic state is described in terms of her attitudes to propositions, then the norms on doxastic states will correspond exactly (in a way to be described later) to the norms given in the body of the paper, provided that the conditions described below are met. If attitudes are individuated by the way they contribute to doxastic states, and if doxastic states are individuated by the norms they are subject to, then the attitudes will *be* the attitudes of belief and lack of belief, regardless of how they are originally described. The language of belief is a kind of universal language for talking about any agent whose doxastic state is properly evaluated in terms of truth or falsity of individual propositions.

The conditions are as follows:

1. The agent's doxastic state can be described by assigning attitudes to some fixed agenda $\mathcal{A}$ of propositions.

2. If p is in $\mathcal{A}$, then $\neg p$ is in $\mathcal{A}$.

3. There are exactly two relevant attitudes, and the agent has exactly one of them to each proposition in $\mathcal{A}$.

4. The evaluation of an agent's doxastic state is given by the sum of scores assigned to attitudes in each proposition — one doxastic state is better than another iff the sum of its scores is higher than the other's.

5. The score of an attitude to a single proposition depends only on which of the two attitudes is held, and whether the proposition is true or false.

6. Neither attitude (strongly or weakly) dominates the other.

7. Attitudes and scores are individuated entirely in terms of the contribution they make to the overall normative status of doxastic states.

B.2 Description

Condition 1 says that an overall doxastic state can be understood in terms of attitudes to individual propositions. The fact that A is allowed to be smaller than the full set of propositions allows for a distinction between propositions on which the agent suspends judgment, and ones towards which she has no attitude whatsoever, as argued in (Friedman, 2012). Conditions 2 and 3 set up further structural features of the situation, with 2 in particular formalizing the idea that an attitude to a proposition implicitly brings with it an attitude to its negation. Condition 4 says how the norms on the state are composed out of norms on individual attitudes. (This condition is considered further in Appendix D.) Condition 5 means that every proposition is scored the same way — no proposition is given more weight than any other, and the relation of an attitude to truth value doesn't depend on the content of the proposition. (Weakenings of this condition are considered in Appendix G, though they undermine some of what is done in this appendix.)

Condition 6, on the other hand, is a plausibility condition rather than being part of the framework. It would be very odd if one of the two attitudes dominated the other — in that case, it's hard to see why anyone would ever have the dominated attitude. Thus, the attitudes would no longer signify anything particularly interesting about the agent's doxastic state. This condition is the same as saying that the two attitudes are scored differently, and that one gets a higher score than the other when the proposition is true, and the other gets a higher score when the proposition is false. Condition 7 makes explicit that the only characterization of doxastic states is structural — it says that these attitudes and values only have functional characterizations, and no intrinsic features that are relevant for the evaluation of doxastic states.

Given the first five conditions, we can call the two attitudes A and B, to remain neutral on what the attitudes actually are. Additionally, we can give labels to the numbers involved in the evaluation. Let us say that the scores for having attitude A to a proposition that is either true or false are, respectively, a_T and a_F. Similarly, the scores for having attitude B to a proposition that is either true or false are, respectively, b_T and b_F.

If we think of the attitudes as just being a tool to describe doxastic states, as suggested by condition 7, then we actually already have some flexibility of description here. Because p is in $\mathcal{A}$ iff $\neg p$ is in $\mathcal{A}$, we can describe new attitudes A' and B' as follows. The agent is said to have attitude A' to p iff she has attitude B to $\neg p$, and she is said to have attitude B' to p iff she has attitude A to $\neg p$. If an attitude doesn't really consist of a mental state that includes the proposition itself, but rather is just some aspect of the overall doxastic state that is subject to two types of evaluation in the two relevant sets of situations,

then since the two sets are complementary, there is no significance to whether an attitude is described as an attitude to p or an attitude to $\neg p$, provided that we keep track of how the evaluation works. If we describe things in terms of attitudes A and B, then the agent is scored by a_T, a_F, b_T, b_F; if we describe things in terms of attitudes A' and B', then the agent is scored by a'_T, a'_F, b'_T, b'_F. But because these two descriptions are supposed to describe the same doxastic state, and having attitude A' to p just *is* having attitude B to $\neg p$, we can see that $a'_T = b_F$, $a'_F = b_T$, $b'_T = a_F$, and $b'_F = a_T$.

B.3 Simplification

Now that we have these eight scores $(a_T, a_F, b_T, b_F, a'_T, a'_F, b'_T, b'_F)$, we can start to say something about how to compare them. Condition 6 says that it is not the case that $a_T \leq b_T$ and $a_F \leq b_F$, and vice versa. In particular, this means that $a_T \neq b_T$ and $a_F \neq b_F$. Thus, either $a_T > b_T$ and $a_F < b_F$, or vice versa. Because the labels 'A' and 'B' are purely arbitrary (by condition 7), we can assume that $a_T > b_T$ and $a_F < b_F$ — if this is not the case, then we should just switch the labels. That is, A is the attitude that is better to have to a true proposition, and B is the attitude that is better to have to a false proposition.

Because $a'_T = b_F$, and so on, we can see that it is also the case that $a'_T > b'_T$ and $a'_F < b'_F$. Thus, for the primed attitudes, again, A' is the attitude that is better to have to a true proposition, and B' is the attitude that is better to have to a false proposition.

Now let us consider how to distinguish the primed attitudes A' and B' from the un-primed attitudes A and B. A state in which the agent has attitude A to a proposition p, and attitude B to its negation $\neg p$ is the same as a state in which the agent has attitude A' to p and B' to $\neg p$. Thus, this is not the case that distinguishes between the primed and un-primed attitudes. Rather, the distinction arises in terms of the evaluation of doxastic states where the agent has the *same* attitude to a proposition and its negation. For the agent to have attitude A to both p and $\neg p$ is for the agent to have attitude B' to both p and $\neg p$, and similarly, to have A' to both p and $\neg p$ is to have B to both p and $\neg p$.

By condition 7, the distinction between the primed and un-primed attitudes will arise in seeing what the normative status is of these cases where an agent has the same attitude to a proposition and its negation. The only relevant normative comparison to be made is whether having A to both p and $\neg p$ is better or worse than having B to both. In the former case, the total contribution to one's overall score is $a_T + a_F$, and in the latter case it is $b_T + b_F$. Thus, the comparison comes down to the question of whether $a_T + a_F$ is greater than, less than, or equal to, $b_T + b_F$.

But recall, $a'_T = b_F$ and $a'_F = b_T$. Thus, $a_T + a_F = b'_T + b'_F$. Similarly, $b_T + b_F = a'_T + a'_F$. Thus, either $a_T + a_F < b_T + b_F$ and $a'_T + a'_F > b'_T + b'_F$, or both inequalities are reversed, or both are equalities. Because the choice of which pair of attitudes to label with primes and which without was purely arbitrary, we can assume that $a_T + a_F \leq b_T + b_F$ — if this is not the case, then we should just switch the labels. That is, for the un-primed attitudes, B is the

better attitude to have, if one has the same attitude to a proposition and its negation, while for the primed attitudes, A' is the better one. Of course, it is always better to have A to the true one of the two and B to the false one, since it is always the case that one is true and the other is false, but it is also always worse to have opposite attitudes to a proposition and its negation if you have A to the false one and B to the true one.

Thus, perhaps after some re-labeling, we have the following conditions:

$$a_T > b_T; a_F < b_F; a'_T > b'_T; a'_F < b'_F$$

$$a_T + a_F \leq b_T + b_F; a'_T + a'_F \geq b'_T + b'_F$$

$$a'_T = b_F; a'_F = b_T; b'_T = a_F; b'_F = a_T$$

B.4 Numerical values

At this point we are ready to start considering what the specific numerical values mean. Condition 7 says that the numerical scores are individuated entirely in terms of the contribution they make to the overall normative status of doxastic states. Additionally, the way a complete doxastic state is evaluated is just by comparing the sum of its scores to the sums of scores of other states. Thus, although it matters whether or not $a_T > b_T$, there are other features of the score that don't matter.

For instance, if one were to add the same value x to both a_T and b_T, this would have no overall impact on the comparison between the total scores of any two doxastic states. This is because we are considering a fixed agenda $\mathcal{A}$ of propositions, and $\mathcal{A}$ always contains a proposition iff it contains the negation of that proposition. If $\mathcal{A}$ has $2n$ propositions in it, then it will always be the case that exactly n of them are true. Replacing a_T by $a_T + x$ and b_T by $b_T + x$ would always entail just adding nx to the total score, no matter what doxastic state is being considered, and no matter which particular propositions are true and which are false. This change doesn't affect the normative status of any doxastic state.

And note — this doesn't affect any of the conditions from earlier either, provided that we also add x to both a'_F and b'_F. We are always either adding the same amount to both sides of an inequality, or leaving both sides unchanged.

Thus, just as we can describe temperature equally well on a Celsius scale and a Kelvin scale, just by adding a constant to all the values, we can rescale the scores by adding a single constant to a_T, b_T, a'_F, b'_F.

Similarly, we can rescale scores by adding a single constant to a_F, b_F, a'_T, b'_T — and there is no need for this constant to be the same or different from the previous one. These two modifications can be made completely independently.

Finally, there is one more numerical transformation that can be made to these scores without changing anything that matters for the overall normative status of doxastic states. In particular, if we multiply or divide all 8 of the numerical values $(a_T, a_F, b_T, b_F, a'_T, a'_F, b'_T, b'_F)$ by the same constant, then we will just multiply or divide the total score of each doxastic state by the same

constant. As long as this constant is positive, this won't change any facts about the ordering, and even if it is negative, nothing will change, provided that we interpret lower scores as better rather than higher scores. This is like changing between a Celsius and Fahrenheit scale for temperature, or turning our thermometers upside down.

Thus, given these three types of numerical transformations that can be applied, we can make some overall simplifications. The choice of which simplification to make will be entirely a matter of convention.

Here is an example, showing that the situation described in the main text is completely general. By adding a single constant to (a_T, b_T, a'_F, b'_F), we can make sure that $b_T = 0$. By adding a single constant to (a_F, b_F, a'_T, b'_T), we can make sure that $b_F = 0$. Thus, the conditions at the end of the previous section simplify to:

$$a_T > 0; a_F < 0$$

$$a_T + a_F \leq 0$$

$$a'_T = 0; a'_F = 0; b'_T = a_F; b'_F = a_T$$

This is exactly the characterization for doxastic states given in the main text, where we interpret A as belief, B as lack of belief, a_T as R, and a_F as $-W$. This leads us to think of B' as disbelief, and A' as lack of disbelief.

The condition that $R < W$ from the main text corresponds to the condition that $a_T + a_F < 0$, which is the choice made earlier in this appendix for how to distinguish the primed attitudes from the un-primed ones. In the main text, this assumption was justified on the basis that believing both a proposition and its negation should be worse than believing neither. After the work of this appendix, it can be argued instead that this is not a significant assumption, but rather a naming convention. The argument that this apparent assumption is not actually significant is the main point of this appendix.

Imagine that we start with one attitude that gets score R if the proposition it is applied to is true, and score $-W$ if the proposition is false, and another attitude that gets score 0 in either case. If $R > W$, contrary to the assumption from the main text, then our conventions from the previous section mean that we should label the first attitude as A' and the second attitude as B'. After rescaling so that $a_T > 0, a_F < 0, b_T = b_F = 0$, we end up with $a_T = W$ and $a_F = -R$. In the characterization given two paragraphs earlier, the attitude we started with (with scores R and W) is interpreted as lack of disbelief, and the other attitude (with constant score 0) is interpreted as disbelief.

The condition that $R < W$ for belief is thus just a convention about whether to call the attitude "belief" or "lack of disbelief". The only substantive assumption it embodies is that $R + W \neq 0$. The only way that the ordering of doxastic states will turn out to be substantively different from the characterization given in the main text is if $R + W = 0$, in which case any coherent state that believes neither p nor $\neg p$ will also have a counterpart that believes both and gets exactly the same score in every situation.

C Three attitudes

In this appendix, I will show that a characterization of doxastic states in terms of belief, disbelief, and suspension will give the same results as the characterization in the main text entirely in terms of belief and lack thereof, given a few conditions.

In particular, I assume that there are only three attitudes (which I will call X, Y, and Z); I assume that these attitudes are evaluated just in terms of truth and falsity as before; I assume that the scores are $x_T, x_F, y_T, y_F, z_T, z_F$ as before; I assume that $x_T > y_T > z_T$ and $x_F < y_F < z_F$, so that none of them dominates either of the others, and so that X is the "positive" attitude, Y is the "neutral" one, and Z is the "negative" one; and most importantly, I assume that the agent has attitude X to p iff she has attitude Z to $\neg p$, so that the positive and negative attitudes are anti-symmetrically distributed among the propositions. I will discuss this symmetry condition further, later in this appendix. (Again, as in the previous appendix, we might allow that these attitudes are only had to an agenda $\mathcal{A}$ that is a subset of the full set of propositions, to represent the distinction between suspending judgment and failing to have an attitude. In that case, I will again require that $\mathcal{A}$ includes the negation of any proposition that it includes.)

As before, the normative statuses of doxastic states will not be changed if we add or subtract a constant k from (x_T, y_T, z_T) or (x_F, y_F, z_F). Thus, by doing this, we can adjust things so that $y_T = y_F = 0$.

For a given pair of propositions $(p, \neg p)$, the symmetry condition means that a doxastic state can only have attitudes (X, Z), (Y, Y), or (Z, X). If p is true, then the scores are, respectively, $x_T + z_F$, 0, and $x_F + z_T$. If p is false, then the scores are, respectively, $x_F + z_T$, 0, and $x_T + z_F$.

But if we consider a two-attitude agent who has $a_T = x_T + z_F$ and $a_F = x_F + z_T$ and $b_T = b_F = 0$, then we see that this two-attitude agent will get exactly the same scores as the original three-attitude agent, by having attitudes (A, B), (B, B), and (B, A). Thus, there is a way to translate any symmetric three-attitude doxastic state into a corresponding two-attitude doxastic state without changing any of the scores. Thus, all the facts about the two-attitude doxastic states carry over to symmetric three-attitude doxastic states.

The symmetry condition, however, seems to violate the ideas that motivated the three-attitude setting though. The point of adding disbelief as a separate attitude from belief is that the distinction between the two was taken to be prior to the concept of negation, so that an agent could conceivably have the concepts of belief and disbelief without thereby disbelieving the negation of everything that she believes, and vice versa.

However, even so, it seems plausible that there is a further requirement that one *ought* to have a doxastic state that satisfies the symmetry condition. Thus, for three-attitude doxastic states, we can define coherence in terms of dominance *together with* the symmetry requirement. Thus, although there will be three-attitude doxastic states that don't correspond to any two-attitude doxastic states, all of the coherent ones will correspond. Thus, although there is some

extra generality gained by moving to the three-attitude situation, this extra generality doesn't change the set of coherent doxastic states.

Allowing violations of the symmetry requirement will implicitly expand the number of attitudes to five rather than three. Just as two attitudes, with no symmetry condition, allow for the distinction between belief, disbelief, and suspension when one considers the pair of attitudes $(A,B),(B,B),(B,A)$ to a proposition and its negation, three attitudes, with no symmetry condition, will allow for a five-way distinction between $(X,Z),(X,Y),(Y,Y),(Y,X),(Z,X)$, or perhaps even seven-way, if (X,Y) is significantly different from (Y,Z) and (Y,X) is significantly different from (Z,Y). If the point is to make a three-way distinction, then we should either restrict to two attitudes, or use three attitudes with a symmetry constraint, and these two frameworks are formally equivalent.

D Numerical vs. ordinal scoring

In Appendix B, I considered different ways of scoring doxastic states, but assumed that they still depended on assigning numerical scores to attitudes based on the truth or falsity of individual propositions, and adding them up to give the overall score of the doxastic state. There are three sorts of worries that are natural to have about this assumption. First, it is not clear what the numerical scores on individual attitudes mean. Second, it is not clear why they should be combined additively, rather than in some non-linear way. Third, the overall score of the doxastic state only enters into normative consideration in an ordinal way (coherence is defined in terms of dominance, which is defined in terms of ordinal comparison of scores, and not precise numerical values), and thus the idea of assigning numerical scores seems suspect.

The first and third worries together suggest a natural alternative approach. If there are n situations, then there are 2^n propositions. Under the original framework, what matters is how many of these propositions the agent believes, and out of those how many of them are true and how many are false. Each of these numbers of right and wrong beliefs must then be placed into an ordinal ranking comparing it to the others. This ranking then implicitly defines the numerical values of the scores, to a certain extent.

If every doxastic state is ranked equally to any state that gets exactly two more propositions right and one more proposition wrong, then this implicitly defines $W = 2R$. If every doxastic state is ranked equally to any state that gets exactly three more propositions right and two more propositions wrong, then this implicitly defines $W = 3R/2$. This ordinal ranking can't precisely define numerical scores if the ratio between R and W is irrational, or is even a fraction whose denominator is larger than 2^n. Thus, if the overall ordinal ranking of doxastic states is fundamental, rather than the numerical scores for being right or wrong, then there could be yet a further level of indeterminacy in the probabilistic representation of a doxastic state, since there is some degree of indeterminacy in the ratio between R and W, and thus in the threshold $W/(R+W)$.

The second concern, about additivity, is a deeper challenge to the framework. It can be mitigated somewhat, since the actual numerical characterization of additivity is unnecessary. But what is necessary is the following set of claims:

- Each set of attitudes has a value, and these values are linearly ordered.
- The value of an attitude depends only on the type of attitude it is, and whether the proposition involved is true or false, and not on which proposition is involved. (This is weakened in Appendix G.)
- There is a function f that takes the values of any two disjoint sets of attitudes and returns the value of their union.
- f is commutative, associative, and monotonic.
- Adding a true belief to any set of attitudes increases the value of the set, and adding a false belief decreases the value of the set. (Slight modifications of this assumption are appropriate if a different set of attitudes is used, as in Appendices B and C.)

If the scoring of sets of attitudes satisfies these conditions, then there is a monotonic function g from real numbers to real numbers such that $g(f(x, y)) = g(x) + g(y)$. Thus, even if the original set of values was not additive, by applying g to it (which just relabels the values) we can make them additive.

I haven't systematically investigated whether there are good justifications for these assumptions, or conversely, whether there are plausible alternatives that violate them. Further investigation of alternative ordinal ranking schemes, and the effects they have on the norm that results from dominance, would be quite helpful here.

E Uniqueness, definiteness, and perturbations of probability functions

In this appendix I will discuss some techniques for constructing additional probability functions that represent a doxastic state, given one probability function that represents it. To do this, I will need to introduce some terminology. Throughout this appendix, I will use "t" for "threshold" as an abbreviation for $W/(R + W)$, which is the important value for representation of doxastic states by probability functions.

Recall that a probability function represents a doxastic state if the doxastic state believes p whenever $P(p) > t$ and fails to believe p whenever $P(p) < t$. Note that the only flexibility in which doxastic state is represented by a proposition comes if $P(p) = t$ for some proposition p. Thus, if P is such that no proposition has probability exactly equal to t, then there is exactly one doxastic state that it represents. In this case, I will say that P is *definite*, and otherwise I will say that P is *indefinite*. I will say that a doxastic state is *representable* iff there is some probability function that represents it, and I will say that

it is *definitely representable* iff there is some definite probability function that represents it.

Now let us consider a relation among probability functions. I will say that P' is an ϵ-*perturbation* of P iff they are defined on the same set of situations, and for every situation s, $0 < |P'(s) - P(s)| < \epsilon$. For any probability function P, and any positive ϵ, there is an ϵ-perturbation P' of P. To see this, just remove less than ϵ probability from some situation that has positive probability on P, and redistribute this probability uniformly among all the situations. Note that this particular perturbation assigns probability 0 only to the empty set. If P is already such that it assigns 0 only to the empty set, and ϵ is the smallest positive value it assigns, then *every* ϵ-perturbation of P also assigns 0 only to the empty set.

Let P be a definite probability function (so that no proposition has probability exactly equal to t). Let δ be the minimum of $|P(p) - t|$ for all p, so that δ is a measure of how close any probability gets to the threshold. (This exists because there are only finitely many propositions.) Let n be the number of situations that P is defined on. Let $\epsilon = \delta/n$. Then any ϵ-perturbation of P will also be definite, and in fact will represent the same unique doxastic state that P does. (Since every situation changes in probability by at most ϵ, and every proposition has at most n situations in it, every proposition changes in probability by at most δ, and thus no proposition can change from having probability greater than t to having probability less than t or vice versa, since δ was defined as the minimum difference between the probability of any proposition and t.) This establishes that any definitely representable doxastic state is in fact represented by infinitely many probability functions that span entire intervals of probabilities for each proposition.

Putting together the results of the last two paragraphs, we can see that if P is definite, then there is some ϵ-perturbation P' of P that represents the same doxastic state, such that $P'(s) > 0$ for all situations s. Putting this together with the **Second Main Result**, we get

> **Third Main Result:** If A is a doxastic state that is represented by a definite probability function P, then A is strongly coherent.

Thus, if A is a doxastic state that is representable, but is not strongly coherent, then any probability function P that represents it must be indefinite (so that there is some proposition p with $P(p) = t$) and must have some situation s such that $P(s) = 0$. By either adding or removing s from p, we can find a distinct proposition that either entails or is entailed by p that also has probability exactly equal to t. I conjecture that the only way a doxastic state can be weakly coherent without being strongly coherent is by believing the one of these that entails the other, but not believing the one that is entailed. If so, then we can see why weakly but not strongly coherent doxastic states are so strange.

One more point. If there are n situations, then there are 2^n propositions, at most half of which can be true or false in any given situation. Thus, the greatest score a doxastic state can conceivably have in a given situation is $2^{n-1}R$ and the lowest score a doxastic state can conceivably have is $-2^{n-1}W$. Thus, if P'

is an ϵ-perturbation of P, then the expected score of A on P' must be similar to the expected score of A on P. There are n terms in the sum for each expected score, and the perturbation can't change the overall expected score by more than $n\epsilon 2^{n-1}(R+W)$ in either direction.

Let P be any probability function. Then some doxastic state that it represents is strongly coherent. To see this, let δ be the minimum difference between the expected scores of a doxastic state that it represents, and any doxastic state not represented by P. Let $\epsilon = \delta/(n2^n(R+W))$. Let P' be any ϵ-perturbation of P. Then by the result of the previous paragraph, any doxastic state represented by P' must have also been represented by P, because nothing else could have passed those in expected score. By earlier results, we can find such a P' such that $P'(s) > 0$ for all s, or such that P' is definite. In either case, any doxastic state represented by P' must be strongly coherent. Since these doxastic states must have already been represented by P, we see that every probability function represents at least one strongly coherent doxastic state, as claimed.

F Non-representable coherent states

In this appendix I describe the counterexamples to the converse of the **Second Main Result**. Recall that it said that if A is a doxastic state that is represented by a probability function P such that every situation has non-zero probability, then A is strongly coherent. Thus, I will discuss doxastic states that are strongly coherent, but for which there is no representing probability function. The smallest such counterexamples involve four situations — as long as there are three or fewer situations, the existence of a representing probability function is necessary, as well as sufficient, for being coherent.

If we ignore permutations of the situations, there are only two doxastic states over four situations that, for particular values of R and W, are strongly coherent and yet not represented by a probability function.

These states are illustrated in Figures 10 and 11. The key for reading these four-situation diagrams is given in Figure 9. Both of these states have the feature that situation 1 is special, while the three other situations are treated symmetrically. Each state is represented for some particular threshold by the probability function with $P(1) = .4$ and $P(2) = P(3) = P(4) = .2$, but the states are counterexample, because they are coherent even for some thresholds for which this probability function (like all others) fails to represent them.

For the state from Figure 10, the overall scores are $(4R, 3R-W, 3R-W, 3R-W)$. For the state from Figure 11, the overall scores are $(7R-W, 5R-3W, 5R-3W, 5R-3W)$. An exhaustive computer search found that as long as $R \leq W$, the state from Figure 10 is always strongly coherent. Similarly, the state from Figure 11 is strongly coherent as long as $R \leq W < 3R$ — if $W \geq 3R$, then it is weakly dominated by the state from Figure 10, but no other state can ever weakly dominate it.

To see that these are counterexamples for certain values of R and W, I have to show that they aren't represented by any probability function for those

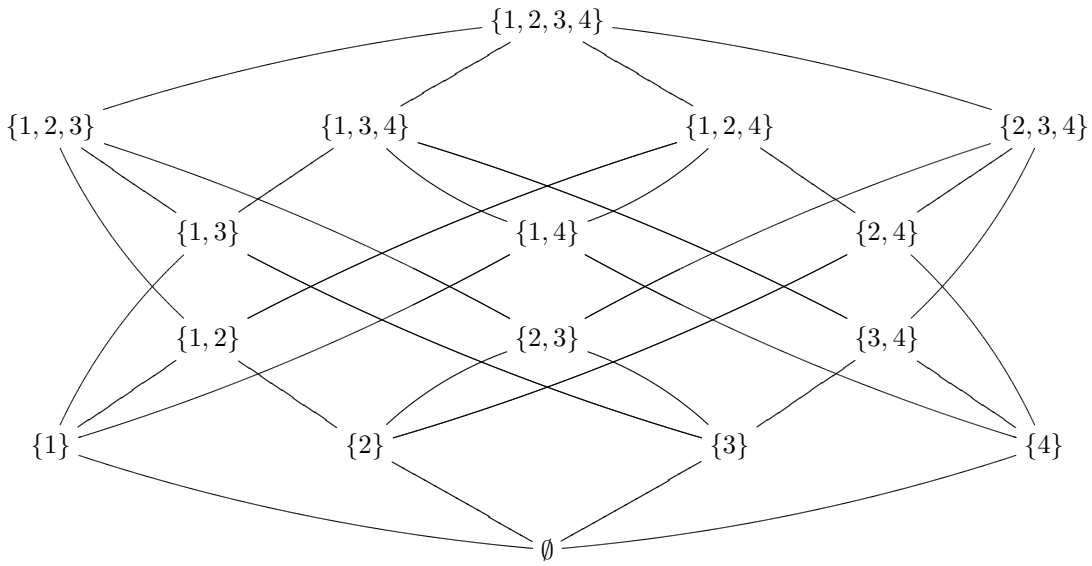


Figure 9: Key for four-situation diagrams.

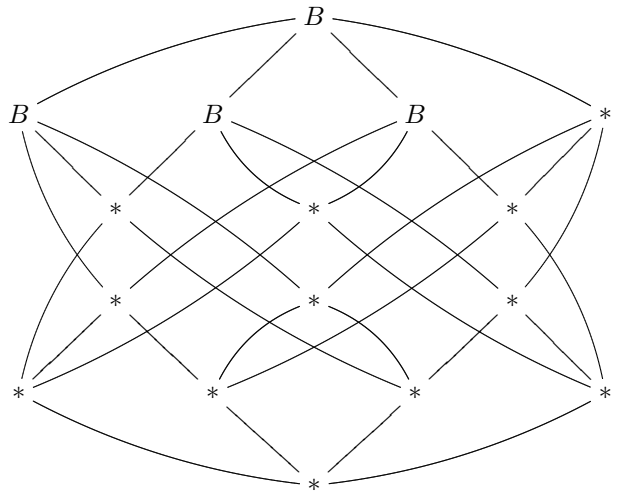


Figure 10: If $R \leq W < 3R/2$, then this doxastic state is coherent but not represented.

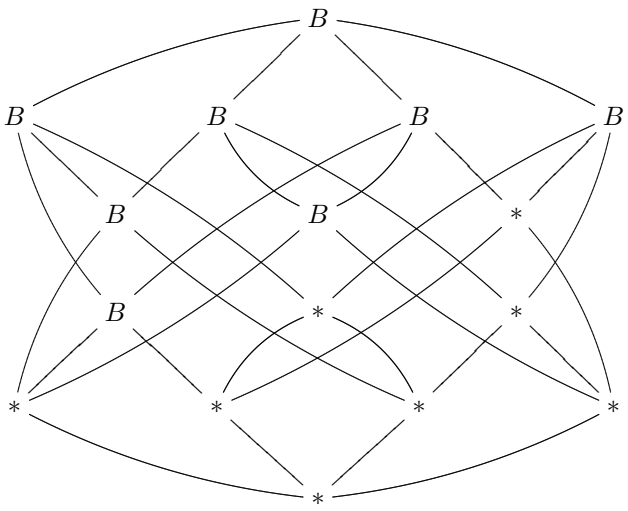


Figure 11: If $3R/2 < W < 3R$, then this doxastic state is coherent but not represented.

values.

Figure 10: If $W \geq 3R/2$, then $W/(W+R) \geq 3/5$. In that case, we can define a probability function P by saying $P(1) = \frac{W-R/2}{W+R}$ and $P(2) = P(3) = P(4) = \frac{R/2}{W+R}$. By some straightforward arithmetic, we can see that the only propositions with probability strictly greater than $\frac{W}{W+R}$ will be $\{1,2,3\}, \{1,3,4\}, \{1,2,4\}$, and $\{1,2,3,4\}$, and thus this probability function would represent the state in Figure 10. However, if $W < 3R/2$, then $W/(W+R) < .6$. In that case, for a probability function to represent this state, it would have to have $P(\{2,3,4\}) < .6$. But by removing whichever one of the three situations has the highest probability, we would see that either $P(\{2,3\}) < .4$ or $P(\{2,4\}) < .4$ or $P(\{3,4\}) < .4$. In any of these cases, the complement of this proposition would have to have probability at least .6, in which case the probability function still fails to represent this doxastic state, because none of those three complements is believed. Thus, this doxastic state is coherent, and yet not represented, iff $R \leq W < 3R/2$.

Figure 11: If $R \leq W \leq 3R/2$ then $1/2 \leq W/(R+W) \leq .6$. In that case, this doxastic state would be represented by the probability function $P(1) = .4$ and $P(2) = P(3) = P(4) = .2$. However, if $3R/2 < W$ then $.6 < W/(R+W)$. In that case, for a probability function to represent this state, it would have to have $P(\{2,3,4\}) > .6$. But by removing whichever one of the three situations has the lowest probability, we would see that either $P(\{2,3\}) > .4$ or $P(\{2,4\}) > .4$ or $P(\{3,4\}) > .4$. In any of these cases, the complement of this proposition would have to have probability less than .6, in which case the probability function still fails to represent this doxastic state, since all three of these complements

are believed. Thus, this doxastic state is coherent, and yet not represented, if $3R/2 < W < 3R$.

Importantly, if $W \geq 3R$, then every strongly coherent doxastic state is represented by a probability function. These counterexamples only work for particular smaller values of W. I haven't yet done an investigation into the counterexamples involving five or more situations. Further such investigations may yield greater understanding of what is going on, which might lead to a plausible strengthening of **Strong Coherence** that does in fact entail probabilistic representability.

G Scoring different propositions differently

One worry about the framework given in the main paper is that it treats every proposition exactly equally. When an agent believes a true proposition she gets the score R regardless of which proposition it is, and when she believes a false proposition she gets the score W regardless of which proposition it is. There are two natural ways to generalize the framework.

G.1 Weighted constant values

The first idea would be to say that propositions might vary in importance, so that each proposition p is accorded a positive weight w_p. Then the contribution that each proposition makes to the overall score of the belief state is multiplied by this weight. Thus, if the agent believes p and it is true, then she gets $w_p \cdot R$, and if the agent believes p and it is false, then she loses $w_p \cdot W$. There are many features of propositions that could conceivably give rise to different levels of importance, and thus affect the values of the w_p, but I will not consider them here.

The first thing to notice is that the **First Main Result** and **Second Main Result** still hold, and exactly the same proofs still work. The overall score of a doxastic state is still the sum of the contributions of each individual proposition, so the expected score is the sum of the expected contributions. The expected contribution of belief in p is now $w_p(R \cdot P(p) - W \cdot P(\neg p))$, and the expected contribution of non-belief is still 0. Thus, belief still gives a higher expected contribution than non-belief if $P(p) > \frac{W}{R+W}$ while non-belief still gives a higher expected contribution if $P(p) < \frac{W}{R+W}$. Thus, despite the different weights, a belief set that is represented by a given probability function has highest expected total score on that probability function, and is therefore not dominated, and is thus still coherent.

For specific weightings, there are some specific belief states that are not represented by any probability function, but are coherent. When the weighting is equal, Appendix F shows how this can arise. But particular weightings can make additional non-represented belief states turn out to be coherent. For instance, consider a case with just 2 situations, so that there are just 4 propositions. As usual, any coherent belief state must believe the proposition that is

true in both situations, and not the one that is not true in either. Thus, there are four doxastic states that could conceivably be coherent.

Two of them are the consistent opinionated ones, in which the agent believes exactly one of the two remaining singleton propositions, and these are always represented by the probability function that assigns 1 to the situation in the believed proposition and 0 to the other, regardless of what R and W are. One of them is the non-committal doxastic state, which believes neither of the two singletons, and which is always represented by the probability function that assigns 1/2 to each of the two situations.

But the remaining inconsistent doxastic state, which believes both singletons, also turns out to be coherent on some weightings of the two propositions, despite never being represented. Regardless of the weightings, it is never dominated by either of the consistent opinionated doxastic states, because it always does better than each in one world and worse in the other. Thus, it is coherent iff it fails to be dominated by the non-committal doxastic state. If we let the weightings of the two singletons be w_1 and w_2, then this doxastic state gets score $w_1 \cdot R - w_2 \cdot W$ in one situation, and $w_2 \cdot R - w_1 \cdot W$ in the other situation. The non-committal state always gets score 0 in each situation. Thus, this inconsistent doxastic state is coherent iff one of these two scores is positive. The first is positive if $\frac{w_1}{w_2} > \frac{W}{R}$, while the second is positive if $\frac{w_1}{w_2} < \frac{R}{W}$. As long as one proposition is weighted sufficiently strongly compared to the other, then there is an additional coherent doxastic state, which turns out not to be represented by any probability function.

The fact that doxastic states that are represented by probability functions are coherent on all weightings, while other doxastic states can be coherent for some weightings but not others, suggests the following conjecture:

> **Main Conjecture:** If A is a doxastic state that is strongly coherent for *every* assignment of positive weights w_p to the propositions, then A is represented by a probability function P.

Some evidence for this conjecture can be given by noting that the doxastic states that are coherent on equal weighting but not represented by any probability function (described in Appendix F) turn out to be incoherent with a different weighting.

Recall that the doxastic state represented in Figure 10 is represented by a probability function as long as $W \geq 3R/2$, but is coherent (with equal weighting) for any $W \geq R$, while the doxastic state represented in Figure 11 is represented if $W \leq 3R/2$, but is coherent (with equal weighting) for any $W < 3R$. For values of W that are close to the critical value of $3R/2$, the respective one of these states that is represented is represented by the probability function that assigns probability 2/5 to the first situation, and 1/5 to each of the others.

If we change the weighting so that all propositions have weight 1, except for the proposition $\{2, 3, 4\}$, which has weight 2, then it turns out that whichever doxastic state is not represented by the relevant probability function is dominated by the other. This is because the difference between the two doxastic

states consists in the fact that the one from Figure 11 believes the three propositions $\{1,2\},\{1,3\},\{1,4\}$ and the double-weighted proposition $\{2,3,4\}$, while the one from Figure 10 believes none of them. In situation 1, the state from Figure 11 gets three additional propositions right and one double-counted proposition wrong, while in the other situations, the state from Figure 11 gets one additional ordinary proposition and one double-counted proposition right, and two propositions wrong. Thus, in every situation, the difference between the scores is $3R - 2W$. If $W > 3R/2$, then the one from Figure 11 is dominated by the one from Figure 10, while if $W < 3R/2$, then the one from Figure 10 is dominated by the one from Figure 11. Allowing this unequal weighting has revealed the incoherence of these doxastic states.

If the conjecture turns out to be right, then we might give a full justification of probabilistic representability by arguing that a doxastic state should be coherent *regardless* of what weights are assigned. Representable doxastic states are coherent regardless of the weights, and the conjecture entails that they are the only ones that are.

G.2 Different values

A further generalization of the main idea of the paper allows the values R and W themselves to change from proposition to proposition. Thus, the overall contribution of a belief in p will be R_p if p is true, and $-W_p$ if p is false. If R_p/W_p is constant, then this is just a notational variant of the previous suggestion, but it is a further generalization if the ratio is allowed to vary. This might be motivated by the thought that some propositions are particularly bad to believe when they are false, while others (perhaps the explanatory ones) are particularly good to believe when they are true. Importantly, the dimension of value should be purely epistemic, and not practical, in order for this to be a proper spelling out of the Truthlove view.

I have not worked out the consequences of this generalization in any great detail, but it prevents one from even stating the original version of the **Main Results**. However, a slightly modified version still holds, with exactly the same proof as before.

> **Modified Main Result:** If there is a probability function P such that doxastic state A believes every proposition p that has $P(p) > \frac{W_p}{W_p+R_p}$ and believes no proposition p that has $P(p) < \frac{W_p}{W_p+R_p}$, then A is coherent.

If the goodness and badness of being right or wrong about different propositions can vary independently, then they each get their own threshold that will play a role in the probabilistic representation.

Many of the results from the main text, like the prohibition on believing a contradictory pair p and $\neg p$, and the requirement to obey single-premise closure, no longer hold in full generality. If $R_p < W_{\neg p}$ and $R_{\neg p} < W_p$, then we still get the requirement that one not believe both p and $\neg p$, but if the values for p and

$\neg p$ can be quite different from one another, then this may not hold. Similarly, if p entails q, then the conditions that require that one not believe p without believing q are that $R_p \leq R_q$ and $W_q \leq W_p$. I have not considered all the motivations that there could be for allowing R and W to differ from proposition to proposition, so it is not clear to me whether these conditions should hold generally.

Of course, the representation may well be farther from being necessary for coherence. And the results from appendices B, C, and D no longer apply, since there are far more ways to score one's attitudes than just by assigning a single value to rightness across all propositions and a single value of wrongness. There is much more room for investigation here.

References

Alchourròn, C. E., Gärdenfors, P., and Makinson, D. (1985). On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50:510–530.

Buchak, L. (2013). *Risk and Rationality*. Oxford University Press.

Christensen, D. (1991). Clever bookies and coherent beliefs. *The Philosophical Review*, 100(2):229–247.

de Finetti, B. (1974). *Theory of Probability*, volume 1. Wiley.

Dreier, J. (1996). Rational preference: Decision theory as a theory of practical rationality. *Theory and Decision*, 40(3):249–276.

Easwaran, K. (2011a). Bayesianism I: Introduction and arguments in favor. *Philosophy Compass*, 6(5):312–320.

Easwaran, K. (2011b). Bayesianism II: Applications and criticisms. *Philosophy Compass*, 6(5):321–332.

Elga, A. (2010). Subjective probabilities should be sharp. *Philosophers' Imprint*, 10(5).

Eriksson, L. and Hájek, A. (2007). What are degrees of belief? *Studia Logica*, 86:185–215.

Foley, R. (1993). *Working Without a Net: A Study of Egocentric Epistemology*. Oxford.

Friedman, J. (2012). Suspended judgment. *Philosophical Studies*.

Gaifman, H. (2004). Reasoning with limited resources and assigning probabilities to arithmetical statements. *Synthese*, 140:97–119.

Garber, D. (1983). Old evidence and logical omniscience in Bayesian confirmation theory. In Earman, J., editor, *Testing Scientific Theories*, volume 10. Minnesota Studies in the Philosophy of Science.

Hacking, I. (1967). Slightly more realistic personal probability. *Philosophy of Science*, 34(4):311–325.

James, W. (1896). The will to believe. *The New World*, 5:327–347.

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press.

Jeffrey, R. (1965). *The Logic of Decision*. McGraw-Hill.

Joyce, J. (1999). *The Foundations of Causal Decision Theory*. Cambridge University Press.

Kolodny, N. (2007). How does coherence matter? *Proceedings of the Aristotelian Society*, 107:229–263.

Leitgeb, H. and Pettigrew, R. (2010a). An objective justification of Bayesianism I: Measuring inaccuracy. *Philosophy of Science*, 77:201–235.

Leitgeb, H. and Pettigrew, R. (2010b). An objective justification of Bayesianism II: The consequences of minimizing inaccuracy. *Philosophy of Science*, 77:236–272.

Lin, H. and Kelly, K. (2012). Propositional reasoning that tracks probabilistic reasoning. *Journal of Philosophical Logic*, 41(6):957–981.

Meacham, C. and Weisberg, J. (2011). Representation theorems and the foundations of decision theory. *Australasian Journal of Philosophy*, 89(4):641–663.

Percival, P. (2002). Epistemic consequentialism. *Proceedings of the Aristotelian Society*, pages 121–151.

Ramsey, F. P. (1926). Truth and probability. In Braithwaite, R. B., editor, *The Foundations of Mathematics and other Logical Essays (1931)*. Harcourt, Brace and Company.

Rumfitt, I. (2000). 'yes' and 'no'. *Mind*, 109:781–823.

Savage, L. J. (1954). *The Foundations of Statistics*. Dover.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall.

Wedgwood, R. (2002). The aim of belief. *Philosophical Perspectives*, 16:267–297.

White, R. (2009). Evidential symmetry and mushy credence. *Oxford Studies in Epistemology*.

Zynda, L. (2000). Representation theorems and realism about degrees of belief. *Philosophy of Science*, 67:45–69.