

# Belief Revision from Probability

Jeremy Goodman

School of Philosophy  
University of Southern California, USA  
jeremy.goodman@usc.edu

Bernhard Salow

Faculty of Philosophy  
University of Oxford, UK  
bernhard.salow@philosophy.ox.ac.uk

In previous work ([5, 6]), we develop a question-relative, probabilistic account of belief. On this account, what someone believes relative to a given question is (i) closed under entailment, (ii) sufficiently probable given their evidence, and (iii) sensitive to the relative probabilities of the answers to the question. Here we explore the implications of this account for the dynamics of belief. We show that the principles it validates are much weaker than those of orthodox theories of belief revision like AGM [1], but still stronger than those valid according to the popular Lockean theory of belief [4], which equates belief with high subjective probability. We then consider a restricted class of models, suitable for many but not all applications, and identify some further natural principles valid on this class. We conclude by arguing that the present framework compares favorably to the rival probabilistic accounts of belief developed by Leitgeb [13, 14] and Lin and Kelly [17].

## 1 Probability Structures

We will work with the following simplification of the models in [5]:

**Definition 1.1.** A *probability structure* is a tuple  $\langle S, \mathcal{E}, Q, Pr, t \rangle$  such that:

1.  $S$  is a non-empty set (of *states*),
2.  $\mathcal{E} \subseteq \mathcal{P}(S) \setminus \{\emptyset\}$  (the *possible bodies of evidence*),
3.  $Q$  (the *question*) is a partition of  $S$ ,
4.  $Pr$  (the *prior*) is a probability distribution over  $S$ , and
5.  $t \in [0, 1]$  (the *threshold*)

Propositions are modeled as subsets of  $S$ , where  $p$  is true in  $s$  if and only if  $s \in p$ . We say that  $E' \in \mathcal{E}$  is the *result of discovering*  $p$  in  $E \in \mathcal{E}$  just in case  $E' = E \cap p$ ; this will allow us to talk about how beliefs evolve in response to changes in one's evidence.

Which propositions an agent believes is a function of their evidence and is also given by a set of states, so that an agent with evidence  $E$  believes  $p$  if and only if  $B(E) \subseteq p$ . This ensures that their beliefs are closed under entailment, and thus already marks a departure from popular ‘Lockean’ accounts of belief [4], according to which one believes a proposition if and only if its probability exceeds a particular threshold. But it is compatible with the more plausible direction of Lockeanism, namely:

THRESHOLD: You believe  $p$  only if  $p$  is sufficiently probable given your evidence.

If  $B(E) \subseteq p$ , then  $Pr(p|E) \geq t$ .

We can think of the members of the question  $Q$  as its *answers*; we write  $[s]_Q$  for the member of  $Q$  containing  $s$ . The proposal in [5] then boils down to claiming that  $s \in B(E)$  if and only if  $s \in E$  and the answers to  $Q$  that are more probable than  $[s]_Q$  have total probability less than the threshold  $t$ . Writing  $Pr_E$  for  $Pr(\cdot|E)$ , this can be formalized as follows:

**Definition 1.2.**  $B(E) = \{s \in E : Pr_E(\{s' : Pr_E([s']_Q) > Pr_E([s]_Q)\}) < t\}$

This means that one believes as much as possible subject to two constraints: (i) THRESHOLD, and (ii) that the totality of one's beliefs corresponds to the conjunction of one's evidence with a disjunction of answers to  $Q$  that includes any answer at least as probable (given one's evidence) as any other it includes. One notable attraction of this proposal is that what one believes corresponds to the discrete analogue of the highest posterior-density region typically used to define 'credible intervals' from probability density functions in Bayesian statistics. A logically significant feature of the proposal, to which we will return later, is that it involves not only local probability comparisons between different answers to  $Q$ , but also a global probability comparison between a collection of such answers and the threshold  $t$ .

## 2 Principles and Results

A core idea behind the orthodox AGM [1] theory of belief revision is that belief revisions are trivial whenever what you learn is compatible with your initial beliefs: you should simply add the discovery to your beliefs, draw out the logical consequences of these beliefs, and leave everything else unchanged. Here we will focus on five principles that encode various aspects of this idea. Exploring when and how these principles can fail will be a useful way of exploring the extent to which our account of belief requires departing from orthodoxy when it comes to belief dynamics. These principles are:<sup>1</sup>

$\Diamond-$  If you don't believe not- $p$  and then discover  $p$ , you shouldn't give up any beliefs.

If  $B(E) \cap p \neq \emptyset$ , then  $B(E \cap p) \subseteq B(E)$ .

$\Diamond R$  If you don't believe not- $p$  and then discover  $p$ , you shouldn't reverse any of your beliefs (i.e. go from believing something to believing its negation).

If  $B(E) \cap p \neq \emptyset$ , then  $B(E) \cap B(E \cap p) \neq \emptyset$ .

$\Box+$  If you believe  $p$  and then discover  $p$ , you shouldn't form any new beliefs.

If  $B(E) \subseteq p$ , then  $B(E) \subseteq B(E \cap p)$ .

$\Box-$  If you believe  $p$  and then discover  $p$ , you shouldn't give up any beliefs.

If  $B(E) \subseteq p$ , then  $B(E \cap p) \subseteq B(E)$

$\Box R$  If you believe  $p$  and then discover  $p$ , you shouldn't reverse any of your beliefs.

If  $B(E) \subseteq p$ , then  $B(E) \cap B(E \cap p) \neq \emptyset$ .

These principles are not logically independent: the  $\Diamond$  principles entail the corresponding  $\Box$  principles, and the  $+$  and  $-$  principles each entail the corresponding  $R$  principles. All of them are valid according to AGM. By contrast, only  $\Box R$  is valid according to Lockean theories that equate believing a proposition with assigning it a sufficiently high probability (for some probability threshold less than 1), and it is valid only if this probability threshold is above  $\frac{\sqrt{5}-1}{2} \approx .62$  (as discussed in [19]).

The present account falls in between these extremes:

**Proposition 1.**  $\Box-$  and  $\Box R$  are valid on the class of probability structures.

**Proposition 2.**  $\Diamond-$ ,  $\Diamond R$ , and  $\Box+$  can all fail in probability structures.

<sup>1</sup>The  $\Diamond$  indicates that the discovery is compatible with your initial beliefs, while the  $\Box$  indicate that it is something you initially believe.  $\Diamond-$  is often referred to as 'preservation'; [19] call  $\Box-$  'weak preservation' and  $\Box R$  'very weak preservation'. If we interpret the non-monotonic consequence relation  $p \sim q$  as saying that  $B(p) \subseteq q$ , then  $\Diamond-$  corresponds to 'rational monotony',  $\Box+$  to 'cut', and  $\Box-$  to 'cautious monotony' in the standard terminology from [12].

We will illustrate Proposition 2 with two examples. Consider a much discussed thought experiment:

### Flipping for Heads

A coin flipper will flip a fair coin until it lands heads.

A natural model of this case is as follows:

$$\begin{aligned} S &= \{s_1, s_2, \dots\} & \mathcal{E} &= \{\{s_i, s_{i+1}, s_{i+2}, \dots\} : s_i \in S\} \\ Q &= \{\{s_i\} : s_i \in S\} & Pr(\{s_i\}) &= \frac{1}{2^i} \\ t &= .99 \end{aligned}$$

Here  $s_i$  is the state in which the coin lands heads on the  $i$ th flip, and  $\{s_i, s_{i+1}, s_{i+2}, \dots\}$  is your evidence if you have just observed the coin land tails on the first  $i - 1$  flips. The question  $Q$  is maximally fine-grained, and the probabilities match the known objective chances.

In this model,  $B(\{s_i, s_{i+1}, s_{i+2}, \dots\}) = \{s_i, s_{i+1}, \dots, s_{i+6}\}$ : you always believe that the coin will land heads within the next seven flips.  $\Diamond-$  is violated whenever you observe the coin land tails. For example, let  $p = \{s_2, s_3, \dots\}$ . Then  $B(S) \cap p \neq \emptyset$ , but  $B(S \cap p) = \{s_2, \dots, s_8\} \not\subseteq \{s_1, \dots, s_7\} = B(S)$ . We think this is exactly the right prediction.

To turn this into a counterexample to  $\Box+$ , we add new body of evidence  $E' = \{s_1, s_2, \dots, s_7\}$  to  $\mathcal{E}$ . Intuitively, we can think of this as the evidence you receive if you walk away from the experiment before the first flip, and are later told that the coin landed heads within the first seven flips. It is easy to verify that  $B(E') = \{s_1, \dots, s_6\}$ . So  $B(S) \not\subseteq B(S \cap E')$ , even though  $B(S) \subseteq E'$ . That this can happen should be unsurprising in a framework like ours in which agents have ‘inductive’ beliefs that go beyond what is strictly entailed by their evidence: discovering something that you previously believed only inductively will strengthen your evidence, putting you in a position to draw further inductive conclusions.

Counterexamples to  $\Diamond R$  are subtler, for reasons we will explain in the next section. But here is one:

### Drawing a Card

You are holding a deck of cards, which is either a fair deck consisting of 52 different cards or a trick deck consisting of 52 Aces of Spades. Your background evidence makes it 90% likely that the deck is fair. You draw a card at random; it is an Ace of Spades.

Here is a possible model of the example:

$$\begin{aligned} S &= \{F_1, F_2, \dots, F_{52}, T\} & \mathcal{E} &= \{S, \{F_1\}, \dots, \{F_{51}\}, \{F_{52}, T\}\} \\ Q &= \{\{F_1, F_2, \dots, F_{52}\}, \{T\}\} & Pr(\{F_i\}) &= \frac{9}{52} \approx .017, Pr(\{T\}) = .1 \\ t &= .85 \end{aligned}$$

The states  $F_i$  are all states in which the deck is fair; they are distinguished only by which card you will draw, with  $F_{52}$  being the one where you draw the Ace of Spades. State  $T$  is the state in which the deck is the trick deck (and you thus draw an Ace of Spades). Your evidence settles all and only what card you drew; so when you draw an Ace of Spades, it leaves open both that you did so by chance and that you did so because it is a trick deck. The question is simply whether the deck is fair. It is easy to see that, according to this model, you should initially believe only that the deck is fair. Your initial beliefs are thus compatible with it being fair and you drawing the Ace of Spades by chance. Yet when you discover that you drew an Ace of Spades, you should reverse your opinion and conclude that you’re holding the trick deck, since  $Pr(\{T\} | \{F_{52}, T\}) \approx \frac{.1}{.1+.017} \approx .855 > t$ .

Note that, in this model, your discovery is not a disjunction of answers to the question  $Q$ . If we changed the question to a more fine-grained one, so that your discovery was a disjunction of its answers, then the case would no longer yield a counterexample to  $\Diamond R$ . For example, relative to the question *is the deck fair and will I draw an Ace of Spades* – i.e. relative to  $Q' = \{\{F_1, F_2, \dots, F_{51}\}, \{F_{52}\}, \{T\}\}$  – you

will initially believe that you won't draw an Ace of Spades, in which case your subsequent discovery isn't compatible with your initial beliefs. And relative to the question *is the deck fair and what will I draw* – i.e. relative to the maximally fine-grained  $Q'' = \{\{s\} : s \in S\}$  – you will initially have no non-trivial beliefs, and in particular you won't start out believing that the deck is fair. In the next section, we will see that this is part of a more general pattern about counterexamples to  $\Diamond R$ .

### 3 Orthogonality

In the previous section, we saw that some of the surprising belief dynamics in probability structures depended on discoveries that cross-cut the question. Notice that structures in which this cannot happen, because every member of  $\mathcal{E}$  is the union of some subset of  $Q$ , satisfy the following constraint:

$$\text{ORTHOGONALITY: } \frac{Pr([s]_Q)}{Pr([s']_Q)} = \frac{Pr([s]_Q|E)}{Pr([s']_Q|E)} \text{ for all } s, s' \in E \in \mathcal{E} \text{ s.t. } Pr([s']_Q|E) > 0$$

This says that the only way that getting new evidence can change the relative probability of two answers to  $Q$  is by completely ruling out one of those answers. While we can ensure ORTHOGONALITY by making the question fine-grained enough to capture all possible discoveries, this isn't always necessary. For example, we could fine-grain the states and bodies of evidence in our model of **Flipping for Heads** to capture the fact that you discover where on the table the coin lands. The bodies of evidence in such a fine-grained model will cross-cut the question *how many time will the coin be flipped*; but, plausibly, ORTHOGONALITY will still hold for this question, since the added information about where the coin lands is probabilistically independent from how many times it will be flipped.

ORTHOGONALITY is interesting because it leads to a stronger logic of belief revision. Firstly,

**Proposition 3.**  $\Diamond R$  is valid on the class of probability structures satisfying ORTHOGONALITY.

Secondly, consider the following principle. It says (roughly) that if you're sure that, whatever you're about to discover, you won't believe a given proposition afterwards, then you already don't believe it:

$\Pi$ – If  $\Pi$  is a partition any member of which you could discover, then there is a  $p \in \Pi$  such that you shouldn't give up any beliefs upon discovering  $p$ .

If  $\Pi \subseteq \mathcal{E}$  is a partition of  $E$ , then  $B(E \cap p) \subseteq B(E)$  for some  $p \in \Pi$ .

We then have the following result:

**Proposition 4.**  $\Pi$ – can fail in probability structures. But it is valid on the class of probability structures satisfying ORTHOGONALITY.

It is also worth noting that  $\Diamond$ – and  $\Box$ – can still fail in structures satisfying ORTHOGONALITY. In particular, ORTHOGONALITY holds in the structures we used in the last section to argue that **Flipping for Heads** yields counterexamples to  $\Diamond$ – and  $\Box$ –.

In our view, a good deal of ordinary talk about what people believe is well-modelled by structures satisfying ORTHOGONALITY. This is because we think that the question  $Q$ , to which attributions of belief are implicitly relativized, typically coincides with the question under discussion in the conversational context in which those attributions are made. Moreover, when a discovery is salient, it is natural to consider a question that is sufficiently fine-grained to capture all the aspects of this discovery that are relevant to its answers. Counterexamples to ORTHOGONALITY (and thus to  $\Diamond R$  and  $\Pi$ –) therefore tend to be 'elusive' in Lewis's [16] sense: attending to these cases often changes the context in such a way that they can no longer be described as counterexamples.

That being said, we do not think that ORTHOGONALITY is plausible as a general constraint. This is because, very often, the only way to ensure ORTHOGONALITY is to adopt a very fine-grained question;

and, often, such fine-grained questions make overly skeptical predictions about what we can believe. Consider, for example, the following case:

### One Hundred Flips

You will flip a fair coin 100 times and watch how it lands each time.

There are natural contexts in which you can be correctly described as initially believing that the coin will not land heads more than 90 times. Our theory predicts this for various natural questions, even for very high thresholds  $t$  – for example, the polar question *will the coin land heads more than 90 times* or the slightly more fine-grained question *how often will the coin land heads*. But neither of these questions satisfies ORTHOGONALITY. For example, discovering that coin lands tails on the first flip will favor ‘no’ over ‘yes’ for the first question, and ‘51’ over ‘49’ for the second question, without ruling out any of these answers. In fact, the only natural question that satisfies ORTHOGONALITY is the maximally fine-grained question *what will the exact sequence of heads and tails be*. But all answers to this question are equally likely, and so this question prevents you from having any non-trivial beliefs about what will happen.

We conclude that ORTHOGONALITY should be rejected as a general constraint, even if it will often hold when we are considering a particular case with a limited range of discoveries.  $\Diamond R$  and  $\Pi-$  are thus not fully general principles of belief revision; but counterexamples are likely to be difficult to pin down.

ORTHOGONALITY is also a fruitful principle in that it helps to facilitate comparisons between our framework and other probabilistic theories of belief. Let us now turn to these.

## 4 Comparisons

In this section, we consider two influential probabilistic accounts from the literature, and compare them with our own account. The first can be seen as a version of our account with an additional constraint imposed on probability structures, and validates  $\Diamond-$  but not  $\Box+$ ; the second can be seen as defining belief from probability structures in a related but different way, and validates  $\Box+$  but not  $\Diamond-$ .

### 4.1 A Stability Theory

The first theory we want to consider is inspired by Leitgeb’s *stability theory of belief* [13, 14]. The guiding idea behind this theory is a probabilistic analogue of  $\Diamond-$  that Leitgeb calls the *Humean thesis*. However, despite the ‘stability’ moniker, the constraints imposed by Leitgeb’s theory are *synchronic* ones relating probabilities, partitions, and thresholds at a single time. So both to facilitate comparison with our framework, and to be (in our view) more faithful to its motivating idea, we will consider a strengthening of Leitgeb’s theory according to which the requirements it imposes on one’s beliefs prior to a discovery continue to hold after one has made that discovery. We can then interpret the view as proposing the following constraint on probability structures:<sup>2</sup>

STABILITY: For all  $E \in \mathcal{E}$  and  $X \subseteq Q$ , if  $Pr(\bigcup X) \geq t$  and  $E \cap \bigcup X \neq \emptyset$ , then  $Pr(\bigcup X|E) \geq t$ .

We then have the following results:

**Proposition 5.**  $\Diamond-$  is valid in probability structures satisfying STABILITY and ORTHOGONALITY. But  $\Box+$  can still fail; and  $\Diamond-$  can fail in structures satisfying STABILITY but not ORTHOGONALITY.

<sup>2</sup>Our STABILITY strengthens Leitgeb’s theory in two ways: first, by identifying the threshold that characterizes the minimal probability of anything one believes with the threshold in terms of which stability is defined, and, second, by not allowing this threshold to be different for different possible bodies of evidence. It also departs from his formulation in quantifying over  $\mathcal{E}$  rather than  $\{\bigcup Y : Y \subseteq Q\}$ ; however, we read him as identifying  $\mathcal{E}$  with  $\{\bigcup Y : \emptyset \neq Y \subseteq Q\}$ , so this is not a substantive departure.

This illustrates how a kind of qualitative stability of belief can be secured by a kind of probabilistic stability (given ORTHOGONALITY), without entailing the full strength of AGM.<sup>3</sup>

We reject STABILITY because we reject  $\Diamond-$  (even in cases where ORTHOGONALITY holds), and along with it the informal idea that rational belief should be stable in anything like the way that Leitgeb claims it should be. STABILITY also places implausible constraints on what agents can believe at a given time. For example, [11] show, in effect, that in **Flipping for Heads** STABILITY entails that the only way to have any non-trivial beliefs about how many times the coin will be flipped is to believe that it will be flipped only once. (This argument depends only on the symmetries of the example, and doesn't depend on whether the coin is fair, biased towards heads, or biased towards tails.) See also [18] and [3].

## 4.2 The Tracking Theory

Lin and Kelly [17] defend a theory which (for reasons we can't explain here) they call the 'tracking theory' of belief. This theory can be seen as an alternative way of defining belief in probability structures, with the parameter  $t$  playing a rather different role. Put informally, a state  $s$  is compatible with your LK-beliefs if there is no answer to  $Q$  that is more than  $\frac{1}{t}$  times more likely than  $[s]_Q$ . Formally:

**Definition 4.1.**  $B_{LK}(E) = \{s \in E : (\forall q \in Q)(Pr_E([s]_Q) \geq t \times Pr_E(q))\}$

In many cases – such as **Flipping for Heads** – the subject will have similar beliefs according to our theory and according to Lin and Kelly's (provided  $t$  is chosen judiciously: low values of  $t$  for Lin and Kelly correspond to high values of  $t$  for us). However, there are important structural differences between the theories. In particular, LK-beliefs are sensitive only to local comparisons of probability between particular answers, while beliefs as we understand them depend also the probabilities of sets of answers. A consequence of this locality is that, as Lin and Kelly note, their theory validates a reasonably strong theory of belief revision (assuming ORTHOGONALITY, which they essentially build in):

**Proposition 6.**  $\Box+$ ,  $\Box-$ ,  $\Box R$ ,  $\Diamond R$ , and  $\Pi-$  are all valid for LK-belief on the class of probability structures satisfying ORTHOGONALITY.

The major shortcoming of the tracking theory, in our view, is that it fails to entail THRESHOLD. Consider a case like **Drawing a Card**, in which one state initially has very low probability (.1) but every other state has even lower probability (.017). Then relative to a fine-grained question such as *is the deck fair and which card will you draw*, you will LK-believe that the deck is a trick deck even for reasonably low values of  $t$  (such as .2). But this belief is only .1 likely on your evidence! And we can, of course, make the case more extreme by increasing the number of distinct cards in the fair deck; so the believed proposition can be arbitrarily improbable for any fixed value of  $t$ .

One might defend the tracking theory against such cases by insisting that we choose a more coarse-grained question; while the theory still fails to entail THRESHOLD, this response at least prevents it from recommending the extreme violations just discussed. However, moving to coarser-grained questions is often in conflict with ORTHOGONALITY. Moreover, the reasons we gave previously for rejecting ORTHOGONALITY as a general constraint applies to the tracking theory as well: just like our theory, the tracking theory will make implausibly skeptical predictions in **One Hundred Flips** unless combined with an ORTHOGONALITY-violating question such as *how many heads will there be*.

Without ORTHOGONALITY, the dynamics of LK-belief are substantially less constrained:

**Proposition 7.**  $\Box+$  and  $\Box-$  are valid for LK-belief on the class probability structures.  $\Diamond R$  and  $\Pi-$  can both fail in such structures.

<sup>3</sup>Leitgeb [14, chapter 4] describes his theory as compatible with AGM (and thus with  $\Box+$ ) since, upon getting new evidence, one may adopt a different, higher threshold than before. But doing so is in no way required by the demands of stability.

$\Box+$  is then the only principle valid for LK-beliefs but not for beliefs as we understand them.

Moreover, without ORTHOGONALITY, the tracking theory invalidates a new principle that holds for belief as we understand it (assuming we restrict to probability structures with  $t > .5$ ). Consider the following variant of **Drawing a Card** (taken from [8], who also makes parallel observations as an objection to Levi's [15] account of belief):

### Drawing a Card v.2

You are holding a deck which could be either a 'fair' deck of 52 different cards, or one of 52 different 'trick' decks that just contain the same card 52 times. Given your background evidence, the probability that you are holding the fair deck is  $\frac{1}{5}$ , with the remaining  $\frac{4}{5}$  distributed evenly across the 52 trick decks. You are about to draw and turn over one card from your deck.

Let us assume that  $Q$  is *which of the 53 possible decks am I holding* and  $t > .25$ . According to the tracking theory, you initially believe that you hold the fair deck, but after drawing a card you believe that you are holding the relevant trick deck. So we have a failure of the following principle, which says (roughly) that if you're sure that, whatever you're about to discover, you'll believe that a given proposition is false, then don't currently believe that the proposition is true:

$\Pi R$  If  $\Pi$  is a partition any member of which you could discover, there is a  $p \in \Pi$  such that you shouldn't reverse any of your beliefs upon discovering  $p$ .

If  $\Pi \subseteq \mathcal{E}$  is a partition of  $E$ , then  $B(E) \cap B(E \cap p) \neq \emptyset$  for some  $p \in \Pi$ .

By contrast, if belief requires probability over a threshold greater than .5 (as it does on our account), this principle cannot fail.<sup>4</sup>

Overall, then, we see few advantages for the tracking theory over our own. Given ORTHOGONALITY, which Lin and Kelly essentially build into their formalism, the tracking theory offers a stronger theory of belief revision. However, the theory violates THRESHOLD, often in dramatic ways. Moreover, to make reasonable predictions in cases like **One Hundred Flips**, both theories need to appeal to coarse-grained questions that conflict with ORTHOGONALITY. Having done so, both theories invalidate many principles of belief revision, although the details differ slightly (with our theory invalidating  $\Box+$  and Lin and Kelly's invalidating  $\Pi R$ ).

## 5 Further work

We conclude with three directions for further work. One concerns nonmonotonic consequence, where  $p \sim q$  is interpreted as  $B(p) \subseteq q$ . We think that distinguishing one's evidence from one's beliefs that go beyond one's evidence offers a productive way of thinking about nonmonotonic consequence, and that the logic resulting from our framework contrasts in interesting ways with the one resulting from Lockean theories of belief (explored in [9]).

The second direction concerns constraints on  $\mathcal{E}$ . Consider, for example, the Monty Hall problem, in which it is crucial that when one gets new evidence about one's environment, one also gets evidence that one has gotten such evidence. We argue in [7] that such cases motivate a *nestedness* requirement on  $\mathcal{E}$ : if two possible bodies of evidence are mutually consistent, then one entails the other. This requirement induce new subtleties in the resulting nonmonotonic logic.

<sup>4</sup>Failures of  $\Pi R$  are to be expected for certain notions of belief that are weaker than the one we are operating with here. For example, your 'best guess' about what deck you are holding plausibly does change no matter what card you draw; and arguably what we 'believe' (in ordinary English) often aligns with our best guesses. See [10] and [2] for discussion.

A third question for future work concerns what happens when probability structures are generalized by making the relevant question a function of one's evidence. [5, Appendix C] motivate this generalization, in order to vindicate certain judgments about a family of examples discussed in [6]. We hope to explore these models in future work; one notable feature is that they invalidate  $\square-$  but still validate  $\Pi R$ .

## Acknowledgements

We thank Kevin Dorst, Josh Pearson, and three anonymous referees for TARK for very helpful comments on earlier versions of this material.

## A Proofs

**Proposition 1.**  $\square-$  and  $\square R$  are valid on the class of probability structures.

*Proof.* Since  $\square-$  entails  $\square R$ , it's sufficient to prove the former. We suppose that  $B(E) \subseteq p$ , and show that  $B(E \cap p) \subseteq B(E)$ .

Note that if  $s \in B(E)$ ,  $[s]_Q \subseteq B(E) \subseteq p$ . So for any  $q \in Q$ , if  $Pr([s]_Q|E) \geq Pr(q|E)$ , then also  $Pr([s]_Q|E \cap p) \geq Pr(q|E \cap p)$ . Contraposing, this means that if  $Pr(q|E \cap p) \geq Pr([s]_Q|E \cap p)$  and  $s \in B(E)$ , then  $Pr(q|E) \geq Pr([s]_Q|E)$ , and so  $q \subseteq B(E)$ .

Moreover, since  $B(E) \subseteq p$ ,  $Pr(B(E)|E \cap p) \geq Pr(B(E)|E) \geq t$ .

Now, note that  $B(E \cap p)$  is the minimal  $X \subseteq E \cap p$  such that (i) if  $s \in X$  and  $Pr(q|E \cap p) \geq Pr([s]_Q|E \cap p)$  for  $q \in Q$ , then  $q \subseteq X$ , and (ii)  $Pr(X|E \cap p) \geq t$ . By the above,  $B(E)$  satisfies both (i) and (ii); so it contains the minimal such  $X$  as a subset. So  $B(E \cap p) \subseteq B(E)$ , as required.  $\square$

**Proposition 2.**  $\diamond-$ ,  $\diamond R$ , and  $\square+$  can all fail in probability structures.

*Proof.* Counter-models are given in the main text.  $\square$

**Proposition 3.**  $\diamond R$  is valid in probability structures satisfying ORTHOGONALITY.

*Proof.* Suppose that  $B(E) \cap p \neq \emptyset$ . Let  $s \in B(E) \cap p$  be such that  $Pr_E([s]_Q) \geq Pr_E([s']_Q)$  for any  $s' \in B(E) \cap p$ . We will show that, given ORTHOGONALITY, there can be no  $q \in Q$  such that  $Pr_{E \cap p}(q) > Pr_{E \cap p}([s]_Q)$ . It follows that  $s \in B(E \cap p)$ , thus establishing  $B(E) \cap B(E \cap p) \neq \emptyset$ .

By ORTHOGONALITY, if  $q \in Q$  and  $Pr_{E \cap p}(q) > Pr_{E \cap p}([s]_Q)$ , then either  $Pr_E(q) > Pr_E([s]_Q)$  or else  $Pr_{E \cap p}([s]_Q) = 0$ . But  $s \in E \cap p$ , so  $Pr_{E \cap p}([s]_Q) \neq 0$ . So suppose  $Pr_E(q) > Pr_E([s]_Q)$ . By the way  $s$  was chosen, it follows that  $q \cap (B(E) \cap p) = \emptyset$ . But  $q \cap p \neq \emptyset$ , since  $Pr_{E \cap p}(q) > 0$ . So  $q \cap B(E) = \emptyset$ . But since  $s \in B(E)$ , this contradicts the assumption that  $Pr_E(q) > Pr_E([s]_Q)$ .  $\square$

**Proposition 4.**  $\Pi-$  can fail in probability structures. But it is valid in probability structures satisfying ORTHOGONALITY.

*Proof.* To see that  $\Pi-$  can fail, consider

$$\begin{aligned} S &= \{s_1, s_2, s_3, s_4, s_5, s_6\} & \mathcal{E} &= \{S, \{s_1, s_3, s_5\}, \{s_2, s_4, s_6\}\} \\ Q &= \{\{s_1, s_2\}, \{s_3, s_4\}, \{s_5\}, \{s_6\}\} & Pr &\text{ is uniform} \\ t &= .65 \end{aligned}$$



Let  $p = \{s_1, s_3, s_5\}$  and  $\Pi = \{p, S \setminus p\}$ . Then  $B(S \cap p) = \{s_1, s_3, s_5\} \not\subseteq \{s_1, s_2, s_3, s_4\} = B(S)$  and  $B(S \cap S \setminus p) = \{s_2, s_4, s_6\} \not\subseteq B(S)$ .

Now suppose  $\langle S, \mathcal{E}, Q, Pr, t \rangle$  satisfies ORTHOGONALITY. To show that  $\Pi-$  holds, we suppose that  $B(E \cap p_i) \not\subseteq B(E)$  for each  $p_i \in \Pi$ , and deduce a contradiction.

For each  $i$ , let  $s_i \in B(E \cap p_i) \setminus B(E)$  be such that  $Pr_{E \cap p_i}([s_i]_Q) \geq Pr_{E \cap p_i}([s]_Q)$  for every  $s \in B(E \cap p_i) \setminus B(E)$ . Since  $s_i \in B(E \cap p_i)$ ,  $Pr_{E \cap p_i}(\{s : Pr_{E \cap p_i}([s]_Q) \geq Pr_{E \cap p_i}([s_i]_Q)\}) < t$ . By ORTHOGONALITY,  $Pr_E([s]_Q) \geq Pr_E([s_i]_Q)$  entails that either  $Pr_{E \cap p_i}([s]_Q) \geq Pr_{E \cap p_i}([s_i]_Q)$  or  $Pr_{E \cap p_i}([s]_Q) = 0$ . So  $Pr_{E \cap p_i}(\{s : Pr_E([s]_Q) \geq Pr_E([s_i]_Q)\}) = Pr_{E \cap p_i}(\{s : Pr_{E \cap p_i}([s]_Q) \geq Pr_{E \cap p_i}([s_i]_Q)\}) < t$ .

Now let  $k$  be such that, for every  $i$ ,  $Pr_E([s_k]_Q) \geq Pr_E([s_i]_Q)$ . Then  $\{s : Pr_E([s]_Q) \geq Pr_E([s_k]_Q)\} \subseteq \{s : Pr_E([s]_Q) \geq Pr_E([s_i]_Q)\}$ , and so  $Pr_{E \cap p_i}(\{s : Pr_E([s]_Q) \geq Pr_E([s_k]_Q)\}) \leq Pr_{E \cap p_i}(\{s : Pr_E([s]_Q) \geq Pr_E([s_i]_Q)\}) < t$  for every  $i$ . But then by the law of total probability,  $Pr_E(\{s : Pr_E([s]_Q) \geq Pr_E([s_k]_Q)\}) < t$ , contradicting the assumption that  $s_k \notin B(E)$ .  $\square$

**Proposition 5.**  $\diamond-$  is valid in probability structures satisfying STABILITY and ORTHOGONALITY; but  $\square+$  can fail in such structures. Moreover,  $\diamond-$  can fail in probability structures satisfying STABILITY in which ORTHOGONALITY fails.

*Proof.* To see that  $\diamond-$  holds, note that  $B(E \cap p)$  is the minimal  $X \subseteq E \cap p$  such that (i) if  $s \in X$  and  $Pr(q|E \cap p) \geq Pr([s]_Q|E \cap p)$  for  $q \in Q$ , then  $q \subseteq X$ , and (ii)  $Pr(X|E \cap p) \geq t$ . Then if  $B(E) \cap p \neq \emptyset$ ,  $Pr(B(E) \cap p|E \cap p) = Pr(B(E)|E \cap p) \geq t$  by STABILITY, so  $B(E) \cap p$  meets condition (ii). Moreover, it meets condition (i) by ORTHOGONALITY. So  $B(E) \cap p$  contains the minimal  $X$  meeting (i) and (ii) as a subset. So  $B(E \cap p) \subseteq B(E) \cap p \subseteq B(E)$ , as required.

To see how  $\square+$  can fail, let  $S = \{a, b, c\}$ ,  $\mathcal{E} = \{S, \{a, b\}\}$ ,  $Q = \{\{s\} : s \in S\}$ ,  $Pr(\{a\}) = .9$ ,  $Pr(\{b\}) = .09$ ,  $Pr(\{c\}) = .01$ , and  $t = .9001$ . This structure satisfies STABILITY.  $\square+$  fails, since  $B(S) = \{a, b\} \not\subseteq B(\{a, b\}) = \{a\}$ .

To see how  $\diamond-$  can fail in the absence of ORTHOGONALITY, consider a probability structure in which  $Q = \{A, B, C\}$ ,  $\mathcal{E} = \{S, E\}$ ,  $Pr(A) = \frac{1}{2}$ ,  $Pr(B) = \frac{1}{4} + \varepsilon$ ,  $Pr(C) = \frac{1}{4} - \varepsilon$ ,  $Pr_E(A) = Pr_E(B) = Pr_E(C) = \frac{1}{3}$ , and  $t = \frac{1}{2} + \varepsilon$ . STABILITY holds, but  $\diamond-$  fails:  $B(S) \cap E \neq \emptyset$ , but  $B(E) = E \not\subseteq B(S) = A \cup B$ .  $\square$

**Proposition 6.**  $\square+$ ,  $\square-$ ,  $\square R$ ,  $\diamond R$ , and  $\Pi-$  are valid for LK-belief on the class of probability structures satisfying ORTHOGONALITY

*Proof.* For  $\diamond R$ , see the proof of Proposition 3; For  $\square+$  and  $\square-$  (and hence  $\square R$ ), see the proof of Proposition 7; for  $\Pi-$ , see [7].  $\square$

**Proposition 7.**  $\square+$  and  $\square-$  are valid for LK-Belief on the class of probability structures.  $\diamond R$  and  $\Pi-$  can both fail in such structures.

*Proof.* The failures of  $\diamond R$  and  $\Pi-$  follow from the failure of  $\Pi R$  described in the main text.

Suppose that  $B_{LK}(E) \subseteq p$ . We will show that  $B_{LK}(E \cap p) = B_{LK}(E)$ , thus establishing  $\square+$  and  $\square-$ .

Since  $B_{LK}(E) \subseteq p$ , we have that if  $s \in B_{LK}(E)$ , then  $[s]_Q \subseteq p$ . So if  $s \in B_{LK}(E)$  then  $\frac{Pr([s]_Q|E \cap p)}{Pr(q|E \cap p)} \geq \frac{Pr([s]_Q|E)}{Pr(q|E)}$  for any  $q \in Q$  such that  $Pr(q|E \cap p) > 0$ . So if  $s \in B_{LK}(E)$ , then for any  $q \in Q$  with  $Pr(q|E \cap p) > 0$ ,  $\frac{Pr([s]_Q|E \cap p)}{Pr(q|E \cap p)} \geq \frac{Pr([s]_Q|E)}{Pr(q|E)} \geq t$ . And if  $Pr(q|E \cap p) = 0$ , then trivially  $Pr([s]_Q|E) \geq t \times Pr(q|E \cap p)$ . So  $s \in B_{LK}(E \cap p)$ .

Moreover, if  $s \notin B_{LK}(E)$ , then  $t > 0$  and there is a  $q \in Q$  such that (i)  $Pr(q|E) \times t > Pr([s]_Q|E)$  and (ii)  $q \cap B_{LK}(E) \neq \emptyset$ . Assuming  $Pr([s]_Q|E \cap p) > 0$  then, by the above,  $\frac{Pr(q|E \cap p)}{Pr([s]_Q|E \cap p)} \geq \frac{Pr(q|E)}{Pr([s]_Q|E)} > \frac{1}{t}$ . In that

case  $s \notin B_{LK}(E \cap p)$ . And if  $Pr([s]_Q|E \cap p) > 0$  then also  $t \times Pr(q|E \cap p) > \times Pr([s]_Q|E \cap p)$ . So in that case also  $s \notin B_{LK}(E \cap p)$ .

So  $B_{LK}(E \cap p) = B_{LK}(E)$ , as required. □

## References

- [1] Carlos Alchourrón, Peter Gärdenfors & David Makinson (1985): *On the Logic of Theory Change: Partial Meet Contraction and Revision Functions*. *Journal of Symbolic Logic* 50, pp. 510–530, doi:10.2307/2274239.
- [2] Kevin Dorst & Matthew Mandelkern (2023): *Good Guesses*. *Philosophy and Phenomenological Research* 105(3), pp. 581–618, doi:10.1111/phpr.12831.
- [3] Igor Douven & Hans Rott (2018): *From Probabilities to Categorical Beliefs: Going Beyond Toy Models*. *Journal of Logic and Computation* 28(6), pp. 1099–1124, doi:10.1093/logcom/exy017.
- [4] Richard Foley (1993): *Working without a net: A study of egocentric epistemology*. Oxford University Press.
- [5] Jeremy Goodman & Bernhard Salow (2021): *Knowledge from Probability*. In Joseph Y. Halpern & Andrés Perea, editors: *Proceedings Eighteenth Conference on Theoretical Aspects of Rationality and Knowledge, TARK 2021, Beijing, China, June 25-27, 2021, EPTCS 335*, pp. 171–186, doi:10.4204/EPTCS.335.15.
- [6] Jeremy Goodman & Bernhard Salow (2023): *Epistemology Normalized*. *Philosophical Review* 132(1), pp. 89–145, doi:10.1215/00318108-10123787.
- [7] Jeremy Goodman & Bernhard Salow (ms): *Belief revision normalized*. Unpublished manuscript.
- [8] Ian Hacking (1967): *Isaac Levi, "Gambling with Truth: An Essay on Induction and the Aims of Science" (Book Review)*. *Synthese* 17, pp. 444–448, doi:10.2307/20114579.
- [9] James Hawthorne (1996): *On the Logic of Nonmonotonic Conditionals and Conditional Probabilities*. *Journal of Philosophical Logic* 25(2), pp. 185–218, doi:10.1007/bf00247003.
- [10] Ben Holguín (2022): *Thinking, Guessing, and Believing*. *Philosophers' Imprint* 22(1), pp. 1–34, doi:10.3998/phimp.2123.
- [11] Kevin Kelly & Hanti Lin (2021): *Beliefs, Probabilities, and Their Coherent Correspondence*. In: *Lotteries, Knowledge, and Rational Belief: Essays on the Lottery Paradox*, Cambridge University Press, pp. 185–222, doi:10.1017/9781108379755.011.
- [12] Sarit Kraus, Daniel Lehmann & Menachem Magidor (1990): *Nonmonotonic reasoning, preferential models, and cumulative logics*. *Artificial Intelligence* 44, pp. 167–207, doi:10.1016/0004-3702(90)90101-5.
- [13] Hannes Leitgeb (2014): *The Stability Theory of Belief*. *Philosophical Review* 123, pp. 131–171, doi:10.1215/00318108-2400575.
- [14] Hannes Leitgeb (2017): *The Stability of Belief: How Rational Belief Coheres with Probability*. Oxford University Press, doi:10.1093/acprof:oso/9780198732631.001.0001.
- [15] Isaac Levi (1967): *Gambling with Truth: An Essay on Induction and the Aims of Science*. MIT Press.
- [16] David Lewis (1996): *Elusive Knowledge*. *Australasian Journal of Philosophy* 74, pp. 549–567, doi:10.1080/00048409612347521.
- [17] Hanti Lin & Kevin Kelly (2012): *Propositional Reasoning That Tracks Probabilistic Reasoning*. *Journal of Philosophical Logic* 41(6), pp. 957–981, doi:10.1007/s10992-012-9237-3.
- [18] Hans Rott (2017): *Stability and scepticism in the modelling of doxastic states: Probabilities and plain beliefs*. *Minds and Machines* 27, pp. 167–197, doi:10.1007/s11023-016-9415-0.
- [19] Ted Shear & Branden Fitelson (2019): *Two Approaches to Belief Revision*. *Erkenntnis* 84(3), pp. 487–518, doi:10.1007/s10670-017-9968-1.