Belief revision normalized[*]

Jeremy Goodman and Bernhard Salow

forthcoming in *Journal of Philosophical Logic*

Abstract

We use the normality framework of Goodman and Salow (2018, 2021, 2023a) to investigate of dynamics of rational belief. The guiding idea is that people are entitled to believe that their circumstances aren't especially abnormal. More precisely, a rational agent's beliefs rule out all and only those possibilities that are either (i) ruled out by their *evidence* or (ii) *sufficiently less normal* than some other possibility not ruled out by their evidence. Working within this framework, we argue that the logic of rational belief revision is much weaker than is usually supposed. We do so by isolating a natural family of orthodox principles about belief revision, describing realistic cases in which these principles seem to fail, and showing how these counterexamples are predicted by independently motivated models of the cases in question. In these models, whether one evidential possibility counts as sufficiently less normal than another is determined by underlying probabilities (together with a contextually determined question). We argue that the resulting probabilistic account of belief compares favorably with other such accounts, including Lockeanism (Foley, 1993), a 'stability' account inspired by Leitgeb (2017), the 'tracking theory' of Lin and Kelly (2012), and the influential precursor of Levi (1967). We show that all of these accounts yield subtly different but similarly heterodox logics of belief revision.

In science and in ordinary life, we rely on beliefs that go beyond what is strictly entailed by our evidence. How should such beliefs evolve in response to new evidence?

Consider the following example:

Bias Detection
You know that a bag is either 'unbiased', containing five red balls and five black balls, or 'red-biased', containing six red balls and four black balls. In fact, it is red-biased. You draw a ball at random from the bag, note its color, put it back in the bag, and repeat.

Unless you have very bad luck, making enough draws from the bag will enable you to know, and hence to rationally believe, that it is red-biased. For if rational

---

inductive belief were impossible here, it is hard to see how it could ever be possible in science, a conclusion we find intolerable.[1]

Assuming you are not very unlucky, then, there will be a first draw from the bag after which your evidence supports believing that it is red-biased; call it draw n. Since draw n tips the scales in favor of believing that the bag is red-biased rather than unbiased, draw n must have been a red ball. Now suppose you draw a black ball on draw $n+1$. This is compatible with what your evidence supported believing after draw n. For even if your evidence had logically entailed that you are drawing from a red-biased bag, the bias is only 60%, which isn't enough to support believing that the next draw isn't going to be a black ball; you should still suspend judgment on the color of the next drawn ball.[2] Moreover, after noting that draw $n + 1$ is a black ball, your evidence no longer supports believing that the bag is red-biased rather than unbiased: you should again suspend judgment. For your evidence didn't support believing that the bag is biased after draw $n - 1$, and drawing one red and one black ball clearly can't tip the scales in favor of believing that the bag *doesn't* contain an equal number of red and black balls.

Suppose that, throughout, you believe all and only what your evidence supports believing. We then have a case where learning something compatible with what you believed (that draw $n+1$ is a black ball) causes you to give up a belief (that the bag is red-biased).

This presents a challenge to the orthodox AGM framework for theorizing about belief revision. According to AGM (Alchourrón et al., 1985), learning something compatible with your beliefs should never lead you to give up any beliefs.[3] This is because AGM embodies a strong kind of conservatism: your beliefs should be consistent and closed under logical consequence, and you should revise your beliefs in the simplest way possible when you can do so while maintaining consistency and closure. So if what you learn is something you already believed, then you shouldn't change your beliefs at all. And if what you learn is something you weren't previously opinionated about, as in **Bias Detection**, then you should end up believing all and only the consequences of what you previously believed together with what you just learned. Only when you learn something surprising – that is, something you previously believed to be false – is the story more complicated. While AGM has not gone unchallenged, we will argue that accommodating cases like **Bias Detection** motivates more radical departures from it than are usually considered, such as sometimes changing your beliefs upon learning something you already believed.

But we agree with AGM that one's beliefs at any given time should be logically consistent and closed under logical consequence: we share its conservative

---

[1] Bacon (2014) and Goodman and Salow (2023a) defend this claim at greater length; Harman (1986, p. 70) and Carter and Hawthorne (forthcoming) discuss similar examples.

[2] This is assuming a notion of belief which corresponds, roughly, to what you have a shot at knowing, as discussed in §3.1.

[3] This is the principle $\Diamond-$ below; see Appendix D.1 for discussion of how it relates to principles formulated in terms of the $*$ operator standardly used to state AGM. Appendix D.2 explains how it, and various other principles we discuss throughout, can be thought of as corresponding to principles from the literature on nonmonotonic logic.

approach to synchronic constraints on rational belief. This means that we cannot vindicate the above verdicts about **Bias Detection** in the most obvious way – by accepting Lockeanism, the view that one should believe all and only the propositions that have high enough probability given one's evidence. This is because Lockeanism recommends having beliefs that are neither consistent nor closed under logical consequence.[4]

This paper shows how the normality framework of Goodman and Salow (2018, 2021, 2023a) can be used to illuminate the dynamics of rational belief in examples like this. We proceed incrementally, offering increasingly flexible models to accommodate different kinds of cases, and seeing which principles of belief revision stand and fall along the way. Although we will argue that few principles of belief revision hold without exception, we will also show that many principles are valid on natural classes of models. For this reason, we recommend a reorientation in theorizing about belief revision: rather than searching for exceptionless generalizations, we should look for productive if not universally appropriate idealizations that have interesting generalizations as consequences.

1   Five principles

Our investigation will focus on the following five principles:[5]

$\Diamond-$ If you don't believe not-p and then learn p, you shouldn't give up any beliefs as a result of learning p.

$\Diamond R$ If you don't believe not-p and then learn p, you shouldn't reverse any of your opinions as a result of learning p.

$\Box+$ If you believe p and then learn p, you shouldn't form any new beliefs as a result of learning p.

$\Box-$ If you believe p and then learn p, you shouldn't give up any beliefs as a result of learning p.

$\Box R$ If you believe p and then learn p, you shouldn't reverse any of your opinions as a result of learning p.

Each of these principles is a consequence of AGM. By contrast, Lockeanism predicts violations of all of these principles except for $\Box R$.[6]

---

[4]This assumes that the Lockean threshold for 'high enough' probability is less than 1, as it must be in order to avoid skepticism in **Bias Detection**. Our verdicts about **Bias Detection** follow from Lockeanism if the threshold for 'high enough' probability is at least .6: the red draw that finally boosts the probability of red bias above the threshold won't push the probability of black on the next draw below the threshold; black on the next draw will then push the probability of red bias back below the threshold.

[5]$\Box/\Diamond$ indicate the prior status of what is learned (whether it was previously believed/not disbelieved), and $+/-/R$ characterize what the principle prohibits (belief gain/loss/reversal). We don't consider the principle $\Diamond+$ because it (absurdly) prohibits gaining new beliefs when you learn things you didn't previously disbelieve.

[6]Counterexamples to $\Box+$ are predicted when p has above-threshold probability and q has

Although we think Lockeanism is mistaken, since it predicts that agents should have inconsistent beliefs, we agree with its prediction that all of the above principles except for $\Box R$ can fail. We will make this case in stages, since in our preferred framework some of these principles are more robust than others, and modelling their failures will require increasingly flexible classes of models. So while we agree with Lockeanism about which principles hold without exception, we disagree about the ubiquity of counterexamples.

Our preferred normality framework shares a certain basic feature with the Lockean approach, which is absent from AGM: a distinction between one's *evidence*, which accumulates monotonically, and one's *inductive beliefs*, which go beyond what one's evidence entails and may evolve in more complicated ways. The need for this kind of distinction is a crucial insight of probabilistic ways of thinking about rational belief, but it is also separable from the specific Lockean account of how probability and belief are related.

2 The normality framework

This section presents a framework for theorizing about belief in terms of the comparative normality of possibilities compatible with one's evidence. After presenting the formalism, we revisit the five principles from the last section and explain how they relate to some natural conditions on normality structures.

A note on terminology: we call these relations of comparative *normality* in deference to the existing literature, but they are meant to be understood through their role in characterizing rational belief and knowledge, not in terms of some pre-theoretical understanding of what is "normal". Indeed, comparative *plausibility* would arguably have been more natural terminology, especially in models where these relations are closely tied to probability (as discussed in §6).

2.1 Normality structures

We will be exploring the dynamics of belief using the following class of structures introduced in Goodman and Salow (2021):

---

above-threshold probability conditional on p but not unconditionally; counterexamples to $\Box-$ are predicted when p has above-threshold probability, and q has above-threshold probability unconditionally but not conditional on p; and counterexamples to $\Diamond R$ (and hence also to $\Diamond-$) are predicted when not-p has below-threshold probability, q has above-threshold probability, and not-q has above-threshold probability conditional on p. Lockeanism entails $\Box R$ if and only if the threshold for 'high enough' probability exceeds $\frac{\sqrt{5}-1}{2} \approx .62$; see Shear and Fitelson (2019) for discussion.

Definition 2.1. A *normality structure* is a tuple $\langle S, \mathcal{E}, W, \succeq, \gg \rangle$ such that:

1. S is a non-empty set (of *states*),

2. $\mathcal{E} \subseteq \mathcal{P}(S) \backslash \{\emptyset\}$ (the *possible bodies of evidence*),

3. $W = \{\langle s, E \rangle : s \in E \in \mathcal{E}\}$ (the set of *situations*),

4. $\succeq$ is a preorder on W (read '$w \succeq v$' as 'w is *at least as normal as v*'),

5. $\gg$ is a well-founded relation on W (read '$w \gg v$' as 'w is *sufficiently more normal than v*'), such that, for any situations w_1, w_2, w_3, w_4:

   (a) If $w_1 \gg w_2$, then $w_1 \succeq w_2$;

   (b) If $w_1 \succeq w_2 \gg w_3 \succeq w_4$, then $w_1 \gg w_4$.

Definition 2.2. We define two functions from situations to sets of situations:

- $R_E(\langle s, E \rangle) = \{\langle s', E' \rangle \in W : E = E'\}$ (*evidential accessibility*)

- $R_B(w) = \{v \in R_E(w) : \forall u \in R_E(w)(u \not\gg v)\}$ (*doxastic accessibility*)

$R_E(w)$ represents the set of situations compatible with your evidence in w, and $R_B(w)$ represents the set of situations compatible with what you believe in w. So understood, the above characterization of R_E encodes the idea that your evidence is *true and transparent* (i.e., your evidence entails what your evidence is), while the characterization of $R_B(w)$ captures the basic idea of the normality-based account of inductive belief: *your beliefs go beyond what is entailed by your evidence insofar as you are entitled to disregard any situation that is sufficiently less normal than any other situation compatible with your evidence.*

The main tradition in doxastic logic, going by back to Hintikka (1962), models belief directly in terms of R_B: an agent believes a proposition in a situation w if and only if that proposition is true in every situation v doxastically accessible from w (i.e., in every $v \in R_B(w)$). This approach is especially useful when we are interested in agents' beliefs about their own beliefs.

But in theories of belief revision, we want a notion of belief such that, after getting new evidence, an agent's new beliefs needn't be inconsistent with their prior beliefs. R_B doesn't do this, since $R_B(w) \cap R_B(v) = \emptyset$ whenever one has different evidence in w and v. We need a more coarse-grained way of modelling agents' beliefs about the world, which brackets their beliefs about what evidence they have. To this end we identify the objects of belief not with propositions (modeled as sets of situations) but as sets of states (which we call *events*). In effect, this allows us to model agents' beliefs *about the (unchanging) state of the world*, while ignoring what they believe about their own (changing) perspective on it.

Our approach will also be coarse-grained in a second way. Rather than treating what one believes as a function of the situation one is in, we will treat it as a function of one's evidence. This is possible because Definitions 2.1-2.2 entail

that what one believes is determined by one's evidence: for all $s, s' \in E \in \mathcal{E}$, $R_B(\langle s, E \rangle) = R_B(\langle s', E \rangle)$.

Summing up: rather than theorizing about belief using the function R_B from situations to sets of situations, we will instead be working with a partial function B from sets of states (the possible bodies of evidence) to subsets thereof (the states compatible with your beliefs given that evidence).

Definition 2.3. $B(E) = \{s' : \langle s', E' \rangle \in R_B(\langle s, E \rangle) \text{ for some } s \in E \text{ and } E' \in \mathcal{E}\}$

Definitions 2.1-2.2 also entail that the only normality relations that make a difference to doxastic accessibility, and hence to belief, are those among situations that agree on your evidence. This makes it possible, and often convenient, to think of $\succeq$ and $\gg$ as evidence-relative relations on states compatible with one's evidence. To this end we adopt the following notational convention:

Convention 2.1. *Let $s \succeq_E s'$ and $s \gg_E s'$ be shorthand for $\langle s, E \rangle \succeq \langle s', E \rangle$ and $\langle s, E \rangle \gg \langle s', E \rangle$, respectively. This shorthand carries the presupposition that $s, s' \in E \in \mathcal{E}$ (just as $B(E)$ presupposes that $E \in \mathcal{E}$).*

Together with Definitions 2.1-2.3, this allows for the following simple and direct characterization of belief:

Corollary 2.1. $B(E) = \{s \in E : \forall s' \in E(s' \not\gg_E s)\}$

Given the above characterization of belief, normality structures might seem unnecessarily complicated. Why not simply take $\gg$ as basic, understood as an evidence-relative relation between states? One reason for the detour through situations and accessibility is to make explicit how belief as we are modeling here relates to belief as it is understood in doxastic logic, and how normality structures relate to more general models in which situations have further structure or in which evidential accessibility is not an equivalence relation; see Goodman and Salow (2023a) and appendices B and C. We include $\succeq$ in addition to $\gg$ because, as we will see below, it allow us to formulate natural principles with important implications for belief dynamics. Additionally, $\succeq$ is needed to model knowledge in addition to belief, as explained Goodman and Salow (2021), and hence to connect theories of belief revision with traditional epistemology, as in Stalnaker (2006) and §3.1 below.[7]

2.2 Belief dynamics in normality structures

Let us now turn to the dynamics of belief. To model these dynamics using normality structures, we treat 'learning' that an event obtains as simply adding the fact that it obtains to your evidence. Since what one believes is a function of one's evidence, we can then investigate belief dynamics by comparing one's beliefs given evidence E to one's beliefs given evidence $E \cap p$, where p is the event that one learns obtains.

---

[7]Dabrowski et al. (1996) is an influential precedent for modelling relata of accessibility relations as, in effect, state/set-of-states pairs.

The five principles from §1 can then be expressed as follows:

$\Diamond{-}$: If $B(E) \cap p \neq \emptyset$, then $B(E \cap p) \subseteq B(E)$.

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

$\Box{+}$: If $B(E) \subseteq p$, then $B(E) \subseteq B(E \cap p)$.

$\Box{-}$: If $B(E) \subseteq p$, then $B(E \cap p) \subseteq B(E)$.

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

We will see shortly that not even $\Box R$ is valid on the class of all normality structures. This fact both illustrates the generality of the normality framework and motivates the search for additional constraints.

To this end, consider the following four potential constraints on comparative normality, which will be helpful in organizing our subsequent investigation:

WEAK STATISM: $\succeq$ is evidence-independent.

If $s \succeq_E s'$, then $s \succeq_{E'} s'$.

STATISM: Both $\succeq$ and $\gg$ are evidence-independent.

If $s \succeq_E s'$, then $s \succeq_{E'} s'$; and if $s \gg_E s'$, then $s \gg_{E'} s'$.

COLLAPSE: All differences in normality are sufficient differences.

If $v \succeq w$ and $w \not\succeq v$, then $v \gg w$.

COMPARABILITY: Situations with the same evidence are related by $\succeq$.

$s \succeq_E s'$ or $s' \succeq_E s$.

How do these constraints interact with our five principles? We can answer this question with a pair of propositions, giving an example of the kind of results we will explore in this paper.

It turns out that each of our five principles can be validated by imposing some combination of the above constraints on comparative normality:[8]

Proposition 1.

- *$\Diamond R$ (and hence also $\Box R$) is valid on the class of normality structures satisfying* WEAK STATISM.

- *Additionally, $\Box{+}$ and $\Box{-}$ are valid on the class of normality structures satisfying* STATISM *(and hence also on the class of structures satisfying both* WEAK STATISM *and* COLLAPSE*)*.

- *Additionally, $\Diamond{-}$ is valid on the class of normality structures satisfying* STATISM, COLLAPSE, *and* COMPARABILITY.

---

[8] All proofs are relegated to Appendix E.

Conversely, there are no further such entailments:

Proposition 2.

1. $\Diamond-$ *can fail in normality structures satisfying* STATISM *and* COLLAPSE, *and also in normality structures satisfying* STATISM *and* COMPARABILITY.

2. $\Box+$ *and* $\Box-$ *can both fail in normality structures satisfying* WEAK STATISM.

3. $\Box R$ *(and hence also* $\Diamond R$*) can fail in normality structures satisfying both* COLLAPSE *and* COMPARABILITY.

Here is a roadmap of the main technical results of this paper. In §4 we describe a potential counterexample to $\Diamond-$ from the literature, which is naturally modelled as trading on failures of COMPARABILITY. In §5 we describe cases in which $\Diamond-$ fails due to failures of COLLAPSE rather than COMPARABILITY, and show that combining STATISM and COMPARABILITY determines a stronger logic of belief revision than combining STATISM and COLLAPSE does. In §6 we introduce a class of normality models with additional probabilistic structure, which satisfy WEAK STATISM and COMPARABILITY but neither STATISM nor COLLAPSE. In §7 we use these models to argue against $\Box+$; notably, these models validate $\Box-$ despite invalidating STATISM. In §8 we motivate a generalization of these models to argue against $\Diamond R$; despite now invalidating even WEAK STATISM, these models still validate $\Box-$. In §9 we prove corresponding results about other probabilistic models of belief revision from the literature. In §10 we describe cases in which $\Box-$ seems to fail, and show how our probabilistic models can be further generalized to accommodate such failures while still validating $\Box R$.

3 On the interpretation of normality structures

This section (which can be skipped without loss of continuity) provides further detail on how the notions of belief, evidence, and learning modelled by normality structures are to be interpreted.

3.1 Evidence, belief, and knowledge

Normality structures were originally introduced to model knowledge. As mentioned at the end of §2.1, these connections to knowledge provide an important anchor in interpreting belief and evidence in normality structures. This section explains the most important of these connections, since they provide a distinctive and opinionated perspective on the question: what is a theory of belief revision a theory *of*?

Knowledge is intermediate in strength between the operative notion of belief and the operative notion of evidence: your evidence is a subset of what you know, which is in turn a subset of what you believe. As with belief, we are interested here in *inductive* knowledge, which typically goes beyond what is logically entailed by your evidence. For example, you can come to know that

you've lost weight by observing the reading of a scale, or come to know that a die is biased by rolling it repeatedly and observing how it lands. Below we will consider more examples like these and show how they can be modeled using normality structures. Throughout we should be understood as modeling *rational* belief; we are not attempting to capture irrational patterns of belief.

Not all (rational) beliefs are knowledge. This is in part because not all such beliefs are true – induction is fallible. But knowledge is the aim of belief: you shouldn't disregard a possibility if, given your evidence, you have no hope of knowing that it does not obtain. The best case for knowledge is when things are most normal: while there are different ways of modeling knowledge in normality structures, they all agree that belief and knowledge coincide in situations that are at least as normal as all other situations in which you have the same evidence; see Goodman and Salow (2023a). These descriptive and normative connections to knowledge distinguish the operative notion of belief from weaker ones (e.g. thinking that a particular horse will win a race) and stronger ones (e.g. Cartesian absolute certainty); see Goodman and Holguín (2023), who suggest that 'being sure' picks out such a notion.

In allowing that we know more than is entailed by our evidence, we are denying that all evidence is knowledge, and hence disagreeing with the letter of the 'E=K' thesis defended by Williamson (2000). But this is not a deep disagreement, since nothing turns on our use of the word 'evidence' to mark the distinction we care about. We would be happy to instead use 'starting points' or some other more neutral terminology. The point of the distinction is best illustrated by its usefulness in modeling particular examples: if what we are interested in is what a rational agent should believe about their weight, or about the bias of a die, then it is usually best to treat as unproblematic their information about the readings of their scale, or about the observed results of die rolls, as well as various bits of background knowledge.

This way of modelling evidence is an idealization. For example, in normality structures one's evidence is *transparent*, in the sense that one's evidence entails what one's evidence is: if $v \in R_E(w)$, then $R_E(v) = R_E(w)$. We agree with Williamson (2000) that this is not true about real agents' evidence, however 'evidence' is understood. Appendix C discusses some strategies for generalizing normality structures to accommodate failures of transparency, but this remains an important area for further work.

One final terminological point. In keeping with the literature on belief revision, we reluctantly use the word "learn" to describe how one's evidence about the state of the world evolves. But what we "learn" in the ordinary sense of the word is what we come to *know*, and coming to know something is neither necessary nor sufficient for it to come to be entailed by your evidence. It isn't necessary because you might have already known the new claim inductively; it isn't sufficient because the new evidence can put you in a position to gain further inductive knowledge. For more on the dynamics of knowledge in the normality framework, and how they depend on the distinction between one's evidence and one's knowledge more generally, see Goodman and Salow (2023a).

3.2 Iterated Belief Revision

A central topic in the post-AGM literature is the search for principles about, and models of, *iterated* belief revision – that is, learning one thing after another. The AGM axioms are silent on this question, and the simple plausibility-order models of AGM revision aren't able to model iterated revision in cases where the propositions that one sequentially learns are jointly inconsistent.[9] Despite the vast and ingenious literature on how to model such cases, our view is that it concerns a pseudo-problem: there are no such cases. This is because learning a proposition entails knowing that it is true; only truths can be known to be true; and no sequence of true propositions is jointly inconsistent.

We share this sensibility with those working in the Bayesian tradition. They typically don't take there to be any special problem of iterated learning: you just keep conditionalizing. *A fortiori*, there is no problem of iterated belief revision for Lockeans, since one's beliefs are determined by one's probabilities which evolve unproblematically. The problem of what to do when you learn p and then learn some q inconsistent with p is not so much solved as rejected: learning must be understood in such a way that this never happens, since one cannot condition $Pr(\cdot|p)$ on any q inconsistent with p.

While normality structures are much more similar to models of AGM than to Lockean models of belief in terms of probabilities, with respect to how they handle iterated belief revision it is the other way around. Only truths can be learned, because one's evidence about the state of the world is always consistent with the state that actually obtains. Of course, it can still happen that as a result of iterated learning you go from believing p to believing q, where these are jointly inconsistent. But this happens even without iterated learning, whenever you learn something that you initially believed to be false.

This feature of our approach helps to resolve interpretive puzzles that arise in applying theories of belief revision to cases in which a series of informants tell conflicting stories. There is debate about whether what you learn should be characterized as what you are told, or that this is what you are told, or both.[10] In the case of conflicting stories, the normality framework rules out the first and third options: those would involve learning something false, since the stories cannot all be true.

One might worry that, on such a conception of learning, what we can learn will be (implausibly) confined to what it is impossible for us to be mistaken about. But this isn't so. The fact that nothing you learn is something that you are mistaken about doesn't mean that nothing you learn is something you could have been mistaken about. What is true is that, in many cases, the propositions that we learn are such that, had we been mistaken about them, we would also have been mistaken about whether we had learned them. For example, assuming that perception is a way of learning about one's environment, misperception of-

---

[9]When these propositions are consistent we could identify the result of learning them sequentially with the result of learning their conjunction.

[10]Rott (2004) endorses the first option; Lin (2019) is sympathetic to the second; Stalnaker (2009) proposes the third.

ten involves mistakenly thinking that one has learned about one's environment. This means having false beliefs about what one's evidence is, and so requires rejecting the idealization that evidence is always transparent, in the sense of settling what your evidence is. This idealization is built into normality structures, and relaxing it is not straightforward. Appendix C explores some strategies for how this might be done. For now, we simply reiterate that the transparency of evidence is a useful idealization in the kind of cases we care about. In **Bias Detection**, for example, we aren't considering the possibility that you might be wrong about the two possible red/black ratios, or that you might be misperceiving or misremembering what colors you've observed on different draws. Such errors are not impossible, and are sometimes important to consider; but they are a distraction when we are trying to understand the structure and dynamics of your inductive beliefs about the red/black ratio in the bag you're drawing from in an ordinary version of the case.

3.3 Learning, hard and soft

Working in a framework in many respects similar to ours, van Benthem (2007) and Baltag and Smets (2008) influentially distinguish between learning 'hard facts' and learning 'soft facts' – *h-learning* and *s-learning*, for short. Cases of h-learning involve irrevocably eliminating all possibilities in which what you've learned is false. By contrast, cases of s-learning involve merely reducing the plausibility of some or all possibilities in which what you've learned is false, so that every possibility in which it is false ends up less plausible than some possibility in which it is true (and not vice versa).[11]

What we have been calling learning in normality structures is like h-learning: it simply eliminates all possibilities incompatible with what you learn. But something reminiscent of s-learning can occur in normality structures that violate STATISM: discovering p can change the comparative normality of states that remain compatible with your evidence (in the sense that $s \succeq_E s'$ and $s \succeq_{E \cap p} s'$ can differ in truth value). However, there is no analogue of *pure* s-learning (s-learning without h-learning), since any change in states' evidence-relative comparative normality must be occasioned by a change in one's evidence.

While we aren't opposed in principle to building more general models that support something like pure s-learning (say by combining the probabilistic models of §6 with a Jeffrey (1965)-style account of probability dynamics), we don't think that the cases discussed below require doing so. Moreover, we think that the example which van Benthem uses to motivate s-learning – namely, that someone smiling at the poker table might allow you to s-learn that they have a good hand – is better modeled as learning (i.e., h-learning) that they smiled and forming an inductive belief that they have a good hand.

---

[11]Soft learning is similar to approaches to iterated belief revision in terms of revising plausibility orders; see Spohn (1988), Boutilier (1996) and Darwiche and Pearl (1997). Compare also Leitgeb and Segerberg (2007), who give models similar to ours in which worlds are modeled as pairs of a state and a plausibility order.

4 Comparability

In this section, we will consider a style of potential counterexample to $\Diamond-$ raised by Stalnaker (1994). We think that such counterexamples, if genuine, turn on failures of COMPARABILITY. We are not convinced that $\Diamond-$ really does fail in these cases, since we are sympathetic to COMPARABILITY for reasons explained in §6. But the cases are still instructive, in part because they suggest a natural minimal departure from the AGM orthodoxy.

The example is as follows:[12]

> **Three Friends**
> Anna knows that Giuseppe, Georges, and Erik are all mono-lingual. On Monday, based on relatively little information, she believes that Giuseppe only speaks Italian, while Georges and Erik only speak French. On Tuesday, she discovers that Giuseppe and Georges actually speak the same language (perhaps by seeing them talk to each other), without discovering what that language is. On Wednesday, she discovers that Giuseppe and Erik also speak the same language, again without discovering what that language is.

It's natural to think that, on Tuesday, Anna should take no view on whether Giuseppe and Georges both speak only Italian or only French, but should still believe that Erik speaks only French. It's also natural to think that, on Wednesday, Anna should give up her belief that Erik speaks only French: at this point, she should no longer take a view on which langugage the three share. If both of these thoughts are correct, **Three Friends** is a counterexample to $\Diamond-$. For it is consistent with what Anna believes on Tuesday that Giuseppe and Erik speak the same language (in particular, that they both speak French); yet when she discovers on Wednesday that they do speak the same language, she gives up her belief that Erik speaks French.

Building on Lin (2019), we can use the following normality structure to accommodate this purported failure of $\Diamond-$:

- S contains eight states. Each state is a triple xyz where $x, y, z \in \{\mathfrak{i}, \mathfrak{f}\}$ respectively indicate Giuseppe, Georges, and Erik's languages.

- Anna's initial evidence is $E_1 = S$; her evidence after learning that Giuseppe and George are co-lingual is $E_2 = \{\mathfrak{iii}, \mathfrak{iif}, \mathfrak{ffi}, \mathfrak{fff}\}$, and her evidence after learning that they're all co-lingual is $E_3 = \{\mathfrak{iii}, \mathfrak{fff}\}$.

- We impose STATISM and COLLAPSE, so the model is fully determined by specifying $\succeq$ as a relation on states. $s \succeq s'$ if and only if the people whose language Anna is initially mistaken about in s are a subset of the people whose language she is initially mistaken about in s'.

---

[12]We have adapted the example slightly, to make it more plausible that the described learning episodes result in the relevant propositions becoming part of ones evidence.

Anne initially believes that Giuseppe speaks Italian and Georges and Erik both speak French: $B(E_1) = \{\text{iff}\}$. But our interest is in what happens when Anne learns that Giuseppe and Erik are co-lingual – i.e., when she learns $p = \{\text{iii}, \text{ifi}, \text{fif}, \text{fff}\}$, so that her evidence changes from E_2 to E_3. Since $B(E_2) = \{\text{iif}, \text{fff}\}$ and $B(E_3) = \{\text{iii}, \text{fff}\}$, $\Diamond-$ fails, as advertised: $p \cap B(E_2) \neq \emptyset$ and $E_3 = E_2 \cap p$, but $B(E_3) \not\subseteq B(E_2)$.

We'd like to make two points about this example. The first is that, although Stalnaker's judgments about the case are not unnatural, they are not unquestionable either. The failure of $\Diamond-$ turns on the fact that, although $\text{iif} \gg \text{iii}$, both are incomparable with fff. But it isn't hard to imagine a basis for comparison: for example, fff, like iif, involves Anne being initially mistaken about only one person while iii involves her being initially mistaken about both Georges and Erik. This suggests an alternative model, which satisfies COMPARABILITY and STATISM and makes iif and fff both equally normal and sufficiently more normal than iii. Such a model would validate $\Diamond-$, with $B(E_3) = \{\text{fff}\} \subseteq B(E_2) = \{\text{iii}, \text{fff}\}$: Anne would always believe that she initially made as few mistakes as possible given her evidence. So we don't think cases like **Three Friends** pose a decisive counterexample to $\Diamond-$ or, more generally, to AGM.[13]

For the second point, which we owe to Ginger Schultheis, consider what would happen according to the above COMPARABILITY-violating model if, after learning that Giuseppe and Georges are co-lingual, Anne had learned instead that Giuseppe and Erik are *not* co-lingual. Let $E_4 = \{\text{iif}, \text{ffi}\}$. Then $E_4 = E_2 \cap (S \backslash p)$, and $B(E_4) = \{\text{iif}, \text{ffi}\}$. So while evidence E_2 supports believing $\{\text{iif}, \text{fff}\}$, this support would be destroyed both by discovering p and by discovering not-p. In other words, although at the intermediate stage Anne believes that Erik speaks French, she would give up this belief if she were either to learn that Giuseppe and Erik are co-lingual or to learn that they are not co-lingual.

This is a surprising prediction: if Anne would give up a belief whatever she might learn about whether Giuseppe and Erik are co-lingual, what business does she have holding that belief now? We can formulate the claim that this cannot happen as follows:[14]

$\Pi-$ If you believe q, then for any finite set of mutually exclusive and jointly exhaustive discoverable events, it is possible to discover one of them while continuing to believe q.

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

Although our main focus in this paper is on the five principles from §1, more complex principles like $\Pi-$ are also useful in exploring different models of belief revision, and we will return to them below.

---

[13]This is not the end of the story: Goodman and Salow (2023a, §5) argue that anti-skeptical considerations about knowledge support a powerful, though not conclusive, argument against COMPARABILITY.

[14]Pearson (ms) discusses a version of this principle under the label 'Anticipation'.

5 Against ◇−

This section considers a different style of counterexample to ◇−. Drawing on Goodman and Salow (2018, 2023a), we describe two such counterexamples, and review how they can be modeled using normality structures violating COLLAPSE while satisfying both STATISM and COMPARABILITY. We will then show that STATISM and COMPARABILITY lead to a strictly stronger theory of belief revision than STATISM and COLLAPSE, as the former but not the latter validates Π−.

The idea motivating COLLAPSE-failures is illustrated by **Bias Detection**. Inductive support is a fine-grained affair, and sufficient evidence for believing an inductive hypothesis often accumulates gradually, in fits and starts. Suppose you have just enough evidence for an inductive belief that the bag is red-biased. For all you believe, the next draw will be a black ball, since even the red-biased bag has 40% black balls. So your beliefs leave open that your incoming evidence will slightly weaken your inductive grounds for believing that the bag is red-biased. Since your inductive support for believing that the bag is red-biased is only barely sufficient, such a weakening will require giving up that belief.

In this section we will focus on two much discussed cases from the literature for which particularly simple normality structures have already been defended. We will return to **Bias Detection** later, since it raises additional complications that go beyond what is required to argue against ◇−.

Here is the first case, from Dorr et al. (2014):

> **Flipping for Heads**
> A coin flipper will flip a fair coin until it lands heads.

We think that, for some number $n > 1$, you should start off believing that the coin will land heads after at most n flips, and that this is all you believe about the outcome of the experiment.[15] If you then see the coin land tails on the first flip, you should now think only that the coin will land heads after at most n more flips – i.e. after $n + 1$ *total* flips. More generally, until the coin lands heads, your beliefs about how many more flips there will be until it lands heads should remain unchanged. In Goodman and Salow (2018) we offered the following normality structure to vindicate these claims. Since STATISM holds in this model, we can specify comparative normality as a relation between states.

$$S = \{s_1, s_2, \ldots\}$$

$$\mathcal{E} = \{E_i = \{s_i, s_{i+1}, s_{i+2}, \ldots\} : i \geq 1\}$$

$$s_i \succeq s_j \text{ iff } i \leq j.$$

$$s_i \gg s_j \text{ iff } i + n \leq j, \text{ for a suitable constant } n > 1.$$

---

[15]This claim is defended by Levi (1996), Hall (1999), Dorr et al. (2014), and Kelly and Lin (2021); for criticism, see Smith (2018a). Hall (1999), Goodman and Salow (2018), and Kelly and Lin (2021) also note that this poses a challenge for ◇−.

In state s_i the coin lands heads on the ith flip, and E_i is your evidence after having observed it land tails $i - 1$ times. As advertised, if your evidence is that the coin is flipped at least i times, then what you believe is that it will be flipped at least i times and fewer than $i + n$ times: $B(E_i) = \{s_i, \ldots, s_{i+n-1}\}$. Now suppose you see the coin lands tails on the first flip. $\lozenge-$ is then violated. Your new evidence is consistent with your prior beliefs, but your new beliefs don't entail your prior beliefs: $B(E_1) \cap E_2 \neq \emptyset$, but $B(E_1 \cap E_2) \nsubseteq B(E_1)$.

Here is the second case, adapted from Goodman and Salow (2023a):

> **Bjorn**
> Wandering through IKEA, having just gorged himself on Swedish meatballs, Bjorn wonders how much he weighs. Luckily, he's in the bathroom section, where two inexpensive digital scales are on display. He weighs himself on both of them.

Here is a simple model of the dynamics of Bjorn's beliefs. Suppose that, prior to weighing himself, he has so little idea how much he weighs that his subsequent beliefs are effectively swamped by the scale's readings. In particular, assume that, for some positive c, what Bjorn will believe upon seeing the first scale read y is that he weighs within c pounds of y. Upon then seeing the second scale read z, what he believes is that he weighs within $\frac{c}{\sqrt{2}}$ ($\approx .71c$) pounds of the average of y and z. That his beliefs are centered on the average of the measurements, and become more precise, is hopefully intuitive; the factor of $\frac{1}{\sqrt{2}}$ is justified by the fact that the standard deviation of the average of n independent samples from a distribution decreases as the squareroot of the number of samples.

These belief dynamics are inconsistent with $\lozenge-$. After stepping on the first scale, Bjorn believes that he weighs between $y - c$ and $y + c$ pounds. Since it is consistent with his beliefs that he weighs $y + c$ pounds, it should also be consistent with his beliefs that the second scale will read $y + c$ pounds. Suppose that happened: Bjorn would then believe that he weighed between $\frac{2y+c}{2} - \frac{c}{\sqrt{2}}$ and $\frac{2y+c}{2} + \frac{c}{\sqrt{2}}$ pounds, giving up his belief that he weighs at most $y + c$ pounds (since $c < \frac{c}{2} + \frac{c}{\sqrt{2}}$) upon learning something consistent with his prior beliefs.

Goodman and Salow (2023a, §6) offers the following model of **Bjorn**, which delivers these results.[16] The model obeys STATISM, allowing us to again specify comparative normality as a relation between states:

$$S = \{\langle x, y, z \rangle : x, y, z \in \mathbb{R}\}$$

$$\mathcal{E} = \{S\} \cup \{E_y : y \in \mathbb{R}\} \cup \{E_{y,z} : y, z \in \mathbb{R}\}$$

$$s \succeq s' \text{ iff } \delta(s) \leq \delta(s')$$

$$s \gg s' \text{ iff } \delta(s) + c^2 < \delta(s')$$

where

$$E_y = \{\langle x', y', z' \rangle \in S : y' = y\},$$

---

[16]We also show how the model can be derived, given natural assumptions, using the probabilistic account of comparative normality discussed below.

$$E_{y,z} = \{\langle x', y', z' \rangle \in S : y' = y \text{ and } z' = z\}, \text{ and}$$

$$\delta(\langle x, y, z \rangle) = (x - y)^2 + (x - z)^2.$$

Here $\langle x, y, z \rangle$ represents the state where Bjorn's weight is x, the first scale reads y, and the second scale reads z. E_y is Bjorn's evidence after seeing the first scale read y, and $E_{y,z}$ is his evidence after then seeing the second read z.

This model vindicates the belief dynamics described above. Before weighing himself, Bjorn has no beliefs about how much he weighs; after weighing himself once and seeing a reading of y pounds, he believes that his weight in pounds is in the interval $[y - c, y + c]$; after weighing himself again and seeing a reading of z, he believes that his weight in pounds lies in the interval $[\frac{y+z}{2} - \frac{c}{\sqrt{2}}, \frac{y+z}{2} + \frac{c}{\sqrt{2}}]$.[17] This leads to failures of $\Diamond-$ whenever the difference between y and z is great enough that the second interval isn't contained in the first, but not so great that Bjorn expected not to see such an extreme disparity. This happens whenever $|y - z| \in (2c - \sqrt{2}c, \sqrt{2}c] \approx (.59c, 1.41c]$.[18]

Although rejecting either of COMPARABILITY and COLLAPSE can lead to failures of $\Diamond-$, the two principles have different implications for the logic of belief revision. For example, we saw at the end of the last section that $\Pi-$ does not hold on the class of normality structures satisfying COLLAPSE and STATISM. By contrast, we have the following result:

Proposition 3. $\Pi-$ *holds on the class of normality structures satisfying* COMPARABILITY *and* STATISM.

In fact, given STATISM, COMPARABILITY yields a strictly stronger theory of belief revision than COLLAPSE does:

Proposition 4. *Any principle that fails in a normality structure satisfying* STATISM *also fails in a normality structure satisfying* STATISM *and* COLLAPSE.

The distinction between cases like **Three Friends**, in which any failures of $\Diamond-$ would be due to failures of COMPARABILITY, and **Flipping for Heads** and **Bjorn**, in which failures of $\Diamond-$ are attributable instead to failures of COLLAPSE, requires theorizing not merely in terms of the relation $\gg$ of one situation being sufficiently more normal than another but also the underlying relation $\succeq$, of one situation being at least as normal as another. Since $\succeq$ constrains $\gg$, principles formulated in terms of it, like COMPARABILITY (and, as we will see below, WEAK

---

[17] $B(S) = \{s' \in S : \delta(s') \le c^2\}$; $B(E_y) = \{s' \in E_y : \delta(s') \le c^2\}$; $B(E_{y,z}) = \{s' \in E_{y,z} : \delta(s') \le \frac{(y-z)^2}{2} + c^2\}$. More generally, if Bjorn were to step on n scales and we were to represent states by $n + 1$-tuples $\langle x, y_1, \ldots, y_n \rangle$, then for $1 \le i \le n$, $B(E_{y_1,\ldots,y_i}) = \{s' \in E_{y_1,\ldots,y_i} : \delta(s') \le \text{TSS}\{y_1, \ldots, y_i\} + c^2\}$, where $\text{TSS}(X) = \Sigma_{x \in X}(x - \overline{X})^2$ and $\overline{X} = \frac{\Sigma X}{|X|}$.

[18] Given natural assumptions, this will be more than a third of the time. Suppose that errors in the scales are independent and are the result of sampling from a Gaussian distribution with standard deviation σ. Then the probability distribution over values of $y - z$ is characterized by a Gaussian distribution with standard deviation $\sigma' = \sqrt{2}\sigma$. Suppose $c \approx 1.96\sigma$ (chosen so that, for all $E \in \mathcal{E}$, Bjorn's beliefs when his evidence is E have $\sim.95$ probability given E of being true). So $(2c - \sqrt{2}c, \sqrt{2}c] \approx (.81\sigma', 1.96\sigma']$. The probability that $|y - z|$ is in this interval is $\sim.37$.

STATISM), can have interesting consequences for the logic of belief revision, even though B is defined only in terms of $\gg$.

6 Normality and Probability

A natural idea in accounting for failures of COLLAPSE is to somehow appeal to probability thresholds. This is because failures of COLLAPSE mean that there are differences in normality that aren't sufficient to warrant inductive belief, and a familiar idea is that increasing probability can eventually make propositions worthy of inductive belief but only if the probability is sufficiently high.

The remainder of the paper develops this basic idea. This section explains how a threshold for what counts as sufficiently probable can be used to generate normality structures, including the models of **Flipping for Heads** and **Bjorn** from the last section. However, we will see in §7 that those examples were in a way misleading, since the probabilistic account, even in its most constrained form, doesn't generally vindicate STATISM. We then motivate some generalizations of the account in §§8 and 10, and argue that it compares favorably to related probabilistic accounts of belief from the literature in §9.

Consider the following class of structures:[19]

Definition 6.1. A *simple probability structure* is a tuple $\langle S, \mathcal{E}, W, Pr, t \rangle$ where:

1. $S, \mathcal{E}, W$ satisfy clauses 1-3 of the definition of normality structures

2. Pr (the *prior*) is a probability distribution over S such that $Pr(E) > 0$ for all $E \in \mathcal{E}$

3. $t \in (0, 1]$ (the *threshold*)

We can use simple probability structures to generate normality structures. Among situations with the same evidence, one is at least as normal as another just in case its component state is at least as probable. (For convenience, we have situations be incomparable in normality whenever they involve different evidence.) To determine when differences in normality constitute sufficient differences, we compare the probabilities of things being at least as abnormal as they are in the relevant situations. Informally, a difference in normality is sufficient if and only if there is sufficiently high probability, conditional on things being at least as abnormal as they are in the more normal situation, that things are still more normal than they are in the less normal situation. This definition is designed to ensure that anything one believes has sufficiently high probability given one's evidence.[20]

---

[19]The structures are a simplification of the more general *probability structures* introduced in Goodman and Salow (2021) and explored in §8 below.

[20]This account of when a difference in normality becomes sufficient can also be combined with non-probabilistic accounts of normality (i.e. of $\succeq$), like the ones discussed by Smith (2016) and Beddor and Pavese (2020). As long as $\succeq$ obeys COMPARABILITY, the account of $\gg$ will still have the desirable consequence that your beliefs must always have above-threshold probability. More generally, the results below depend only on $\succeq$ obeying COMPARABILITY and, where relevant, WEAK STATISM.

Definition 6.2. $\langle S, \mathcal{E}, W, \succeq, \gg \rangle$ is *determined by* a simple probability structure $\langle S, \mathcal{E}, W, Pr, t \rangle$ if and only if:

- $s \succeq_E s'$ iff $Pr(\{s\}|E) \geq Pr(\{s'\}|E)$;

- $s \gg_E s'$ iff $Pr(\{s'' : s' \not\succeq_E s''\}|\{s'' : s \succeq_E s''\}) \geq t$;

- situations involving different evidence are not related by $\succeq$ or $\gg$.

Proposition 5. *If $\langle S, \mathcal{E}, W, \succeq, \gg \rangle$ is determined by $\langle S, \mathcal{E}, W, Pr, t \rangle$, then it is a normality structure, and $B(E)$ is always the minimal subset of E that*

1. *includes the most probable members of E,*

2. *includes all members of E at least as probable as any it contains, and*

3. *has probability at least t conditional on E.*

More exactly, $B(E) = \left\{ s \in E : Pr_E\left(\left\{ s' : Pr_E(\{s'\}) > Pr_E(\{s\}) \right\} \right) < t \right\}$.

This direct characterization of belief in terms of probabilities is equivalent to the one proposed by Cantwell and Rott (2019) and investigated further by Wang (2022). It can also be straightforwardly generalized to continuous distributions and probability density functions (needed to handle examples like **Bjorn**); see Goodman and Salow (2021, appendix B). So generalized, the analogue of Proposition 5 is then that what one believes is given by the high posterior density region for a given threshold, which is the standard way in Bayesian statistics of summarizing probability distributions using regions; see Kruschke (2014). This correspondence is one reason why we think the above is a more attractive picture of how belief is related to probability than the Lockean picture mentioned earlier and discussed further below.

We can also reformulate this proposal in terms of degrees of normality. Let a situation's *degree of normality* be the probability, given its evidence, that things are at least as abnormal as they are in that situation (so that, for example, a situation has degree of normality 1 whenever it is at least as normal as every other situation with the same evidence). In normality structures determined by simple probability structures, we can directly characterize $\succeq$ and $\gg$ in terms of degrees of normality:

Proposition 6. *If $\langle S, \mathcal{E}, W, \succeq, \gg \rangle$ is determined by $\langle S, \mathcal{E}, W, Pr, t \rangle$, then*

1. *$s \succeq_E s'$ iff $d(\langle s, E \rangle) \geq d(\langle s', E \rangle)$; and*

2. *$s \gg_E s'$ iff $1 - \frac{d(\langle s', E \rangle)}{d(\langle s, E \rangle)} \geq t$.*

where $d(\langle s, E \rangle) = Pr(\{s' : s \succeq_E s'\}|E)$.

Remark 6.1. It is possible to generate normality structures from simple probability structures that agree with Definition 6.2 on $\succeq_E$ and $\gg_E$ but in which all situations are comparable, even situations with different evidence: simply let

$w \succeq v$ iff $d(w) \geq d(v)$, and $w \gg v$ iff $1 - \frac{d(v)}{d(w)} \geq t$.[21] These relations satisfy the definition of a normality structure.[22]

Simple probability structures can also be used to motivate the normality structures proposed in §5. In particular, the structure used to model **Flipping for Heads** is determined by the obvious simple probability structure for the case (in which each state has a probability equal to its initial chance of obtaining).

Normality structures generated in this way satisfy many of the conditions we have discussed, and determine a fairly strong logic of belief revision:

Proposition 7. *Every normality structure determined by a simple probability structure satisfies* COMPARABILITY *and* WEAK STATISM, *so* $\Diamond R$ *and* $\Box R$ *are valid.* $\Box-$ *and* $\Pi-$ *are also valid.*

But not all normality structures generated in this way are as well behaved as the models of **Flipping for Heads** and **Bjorn** discussed above:

Proposition 8. STATISM *and* $\Box+$ *can both fail in normality structures determined by simple probability structures.*

In the next section we argue that such failures of STATISM are an advantage rather than a deficiency of probabilistically generated normality structures.

7 Against $\Box+$

Here, in outline, is our case against STATISM: there are cases where $\Diamond-$ fails and the best explanation of these failures is that COLLAPSE fails too; the best account of how COLLAPSE fails in these cases is that comparative normality aligns with underlying probabilistic structure in the way described above; this probabilistic account of comparative normality will not in general deliver STATISM. In this section we make this argument more concrete, by showing how STATISM fails in **Flipping for Heads** if we expand the possible bodies of evidence. These failures of STATISM also lead to failures of $\Box+$; we defend this prediction.

We begin with an informal statement of the argument. Suppose that instead of watching the coin, you ask someone after the fact whether the coin landed heads in the first n flips (which is the strongest thing you initially believed), and you discover that it did. Although you believed this already, it was not previously entailed by your evidence, whereas it now is. Since your evidence has been strengthened, you may now be in a position to form new inductive beliefs. In particular, you may now believe that the coin landed heads in the first $n-1$ flips. The underlying probabilistic fact is that this proposition can have above-threshold probability given your stronger evidence despite having below-threshold probability given your original evidence. When comparative normality

---

[21]Dorr et al. (2023) argue that all gradable expressions in natural language obey this strong form of comparability.

[22]Goodman and Salow (2021, note 8) claimed that this was impossible, apparently assuming that a situation's degree of normality would have to be its probability considered in isolation.

is then analyzed in terms of simple probability structures, this leads to a failure of STATISM: although, relative to your original evidence, the coin landing heads on the nth flip isn't sufficiently less normal than it landing heads on the first flip, it is sufficiently less normal relative to your new, stronger evidence.

Here is the argument more formally. We modify the relevant probability structure by adding $E' = \{s_1, \ldots, s_n\}$ to $\mathcal{E}$. Now suppose $t = 1 - .5^{n-1} + \epsilon$. Then $B(S) = E'$ and $B(E') = \{s_1, \ldots, s_{n-1}\}$. So $\square+$ fails: $B(S) \nsubseteq B(B(S))$. This is possible because STATISM fails too: $s_1 \nsucceq_S s_n$, but $s_1 \succeq_{E'} s_n$.[23]

This example suggests that STATISM is the exception rather than the rule. It holds in our earlier probablistically derivable models of **Flipping for Heads** and **Bjorn** only because the symmetries of the probability distributions involved and the circumscribed possible bodies of evidence conspire in just the right way. On reflection, we think that failures of $\square+$ should be expected on any framework which, like ours, distinguishes one's evidence from one's beliefs. This is because learning what one previously believed inductively is a way of strengthening one's evidence, so we should expect it to sometimes yield new beliefs. And this requires failures of STATISM.[24]

Since the account predicts failures of $\square+$, it also predicts failures of principles which entail it. Consider the widely endorsed idea that an inductive hypotheses q is only reasonable to believe in response to learning p if, before that discovery, it was reasonable to believe the corresponding material conditional $p \supset q$:[25]

> FRONTLOADING
> If you would believe q after learning p, then you already believe $p \supset q$.
>
> $B(E) \cap p \subseteq B(E \cap p)$.

This principle has considerable intuitive appeal, especially for those in a probabilistic frame of mind. For example, White (2006) influentially emphasizes that the conditional probability of a material conditional given its antecedent is typically lower and never greater than its unconditional probability; it seems surprising, then, that learning its antecedent could make the conditional reasonable to believe when it wasn't beforehand.

Despite its intuitive appeal, we are committed to rejecting FRONTLOADING, since it entails $\square+$. Again, we think that on reflection this is a natural result. Consider how FRONTLOADING fails in our model of **Flipping for Heads**: you come to believe the material conditional *the coin is flipped at most n times $\supset$ the coin is flipped at most $n-1$ times* as a result of learning its antecedent. This conditional is false only in state s_n. While it is true that the probability that s_n

---

[23]This argument works for most $t \in (1 - .5^{n-1}, 1 - .5^n]$, but not for t very close to $1 - .5^n$.

[24]One complication: it might be argued that any purported failure of $\square+$ (or $\square-$) can be re-described in a more fine-grained way where the principle vacuously holds, since usually what we learn isn't something that we anticipated in complete detail (e.g., where on the table a coin lands). Note that counterexamples to principles like FRONTLOADING and $\Pi+$ (discussed below) are unaffected by such fine-grained re-description.

[25]The name is from Chalmers (2012), who defends a closely related principle about knowledge; Hawthorne (2002) and Bacon (2014) also defend analogous principles about knowledge, which Goodman and Salow (2023a) discusses under the heading INDUCTIVE ANTI-DOGMATISM.

obtains is higher after your discovery, so too is the probability that things are more normal than they are in s_n. This is why, on our view, you should come to believe that s_n does not obtain.[26]

FRONTLOADING also entails the analogue of $\Pi-$ for belief acquisition:[27]

$\Pi+$ If you don't believe q, then for any finite set of mutually exclusive and jointly exhaustive discoverable events, it is possible to discover one of them without coming to believe q.

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

This principle says that if you would believe q after learning the answer to a given question regardless of what the answer was, then you should already believe q. It too fails in **Flipping for Heads**. Having added $E' = \{s_1, \ldots, s_n\}$ to $\mathcal{E}$, we should do the same for $S \backslash E' = \{s_{n+1}, \ldots\}$, since you could have been informed that the coin was flipped more than n times. These two mutually exclusive and jointly exhaustive bodies of evidence both support believing that s_n does not obtain (the former inductively, and the latter deductively); but trivial evidence S does not support this belief, in violation of $\Pi+$.

Summing up: normality structures determined by simple probability structures are more friendly to orthodox principles of belief preservation than they are to structurally parallel orthodox principles of belief acquisition. Although $\Pi+$ and even $\Box+$ can fail in such structures, $\Box-$ and even $\Pi-$ hold. This is notable because all of these principles can fail in normality structures that merely satisfy COMPARABILITY and WEAK STATISM.

8   Question sensitivity, and how $\Diamond R$ can fail

There is a problem with normality structures determined by simple probability structures. On the one hand, we need to model states in a way that is sufficiently fine-grained to capture all relevant features of an agent's evidence. On the other hand, there are certain inductive beliefs that we think ordinary agents have. The problem is that these two requirements can conflict with each other.

---

[26]By the same token, these models allow for failures of $\Diamond-$ where what you learn raises the probability of the belief you lose. Imagine rolling a 100-sided die until it lands 1. As in our model of **Flipping for Heads**, we identify states with the number of rolls that this takes. For $t = .9$, $B(S) = \{1, \ldots, 230\}$ (since $1 - .99^{229} < .9 < 1 - .99^{229}$). Subsequently learning $\{230, 231\} \cup \{232, 234, 236, \ldots\}$ would destroy your belief in $\{1, \ldots, 230\} \cup \{232, 234, 236, \ldots\}$ despite increasing its probability from $\approx.95$ to $\approx.99$.

[27]Conversely, $\Pi+$ entails FRONTLOADING whenever $\mathcal{E}$ is *partition closed*, in the sense that, if $E, E' \in \mathcal{E}$ and $E' \subseteq E$, then $E' \in \Pi$ for some $\Pi \subseteq \mathcal{E}$ that is a finite partition of E. Partition closure is a natural constraint on normality structures (and one that is consistent with our argument in appendix A against the closure of $\mathcal{E}$ under non-empty subsets).

Here is an example that illustrates the tension:

Computer Number
You have very strong evidence that your computer is working: the probability that it will turn on is .99999 given your evidence. If it turns on it will display a random ten-digit number.

If the computer turns on, you'll discover what number it displays. So states should be fine-grained enough to take a stand on what number is displayed. It follows that states in which the computer turns on cannot have probability above 10^{-10}. But for all we have said there is only one state in which the computer doesn't turn on and the monitor remains blank, and it has probability 10^{-5}. This would then be the most normal state given your evidence in the normality structure determined by such a probability structure. This yields the intuitively incorrect prediction that you don't believe that the computer will turn on.[28]

The moral of such cases is that, in determining comparative normality, we need to be able to abstract away from many of states' fine details. We can implement this idea by having comparative normality be determined not by the probabilities of particular states, but by the probabilities of equivalence classes of states. This equivalence relation corresponds to a contextually determined question: two states are equivalent just in case they agree on the answer to the question. More precisely:[29]

Definition 8.1. A *probability structure* is a tuple $\langle S, \mathcal{E}, W, Q, Pr, t \rangle$ where Q (the *question*) is a partition of S, and $S, \mathcal{E}, W, Pr, t$ are as in Definition 6.2.

Probability structures determine normality structures in exactly the same way that simple probability structures do, except with the probability of $[s]_Q$ in place of the probability of s in the definition of $\succeq_E$, where $[s]_Q$ is the cell of Q containing s. More precisely, we adopt the following analogue of Definition 6.2 which yields an appropriate analogue of Proposition 5:

Definition 8.2. $\langle S, \mathcal{E}, W, \succeq, \gg \rangle$ is *determined by* the probability structure $\langle S, \mathcal{E}, W, Q, Pr, t \rangle$ if and only if:

- $s \succeq_E s'$ iff $Pr([s]_Q | E) \geq Pr([s']_Q | E)$;

- $\gg_E$ is as in Definition 6.2;

- situations involving different evidence are not related by $\succeq$ or $\gg$.

---

[28]Compare Hacking (1967), who raises a similar case as a challenge for Levi (1967).

[29]This definition is from Goodman and Salow (2021), except for the requirement that $t > 0$, which we failed to notice is needed to ensure that $\gg$ is asymmetric; Hong (2023) independently proposes the same account. (We could simulate $t = 0$ by existentially quantifying over all positive thresholds in the clause for $\gg$.)

Proposition 9. *If $\langle S, \mathcal{E}, W, \succeq, \gg \rangle$ is determined by $\langle S, \mathcal{E}, W, Q, Pr, t \rangle$, then it is a normality structure in which $B(E)$ is always the minimal subset of E that*

1. *overlaps the members of Q most probable given E,*

2. *overlaps all members of Q at least as probable, given E, as any it overlaps,*

3. *contains every state in E that agrees on the answer to Q with any other state it contains, and*

4. *has probability at least t conditional on E.*

More exactly, $B(E) = \left\{ s \in E : Pr_E\left(\left\{ s' : Pr_E([s']_Q) > Pr_E([s]_Q) \right\} \right) < t \right\}.$

Strictly speaking, simple probability structures are not probability structures. But, in a slight abuse of notation, we will speak as if they are, since adding the question $Q = \{\{s\} : s \in S\}$ to a simple probability structure yields a probability structure that determines the same normality structure.

The most important way in which probability structures generalize simple probability structures is that the normality structures they determine need not satisfy WEAK STATISM. This is because the comparative probability of $[s]_Q$ and $[s']_Q$ can change as a result of getting new evidence compatible with both s and s'. For this reason, it will be useful to isolate the following principle, which characterizes the class of probability structures in which this never happens:

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

A special case of probability structures satisfying this condition are those in which every member of $\mathcal{E}$ is a union of some set of cells of Q (which trivially includes simple probability structures). But even in cases where the question we care about is more coarse grained than the discoveries we can make, ORTHOGONALITY is still often a natural idealization. For example, we could fine-grain our model of **Bjorn** to allow Bjorn to discover the font of the digital scales' display; this requires fine-graining states, but in a way that is plausibly orthogonal to the question *what does Bjorn weigh and what will the scales read.* This also illustrates the fact that, when ORTHOGONALITY holds and we are only interested in beliefs about events that are unions of cells of Q, it is harmless to use simple probability structures, where Q is eliminated as an additional parameter and instead simply identified with S (with the obvious adjustments).

If we restrict our attention to probability structures that satisfy ORTHOGONALITY, we have the following analogue of Proposition 7:

Proposition 10. *Normality structures determined by probability structures that satisfy ORTHOGONALITY satisfy WEAK STATISM, so $\Diamond R$ and $\Box R$ are valid. $\Box-$ and $\Pi-$ are also valid.*

However, when we consider the full range of probability structures, more principles can fail, although the resulting class of normality structures still validates some non-trivial principles:

Proposition 11. *Every normality structure determined by a probability structure satisfies* COMPARABILITY *and validates* $\Box-$ *and* $\Box R$.

Proposition 12. WEAK STATISM, $\Diamond R$, *and* $\Pi-$ *can all fail in normality structures determined by probability structures.*

Failures of $\Diamond R$ are easily illustrated by our opening example of **Bias Detection** where Q is the polar (i.e. two-cell) question *is the bag red-biased*. For any $t < 1$, there will be a pair of sequences of red and black observations, the second of which is a continuation of the first, such that the probability that the bag is red-biased is above t conditional on the first sequence and the probability that the bag is unbiased is above t conditional on the longer second sequence. Since, relative to this question, all situations compatible with your evidence in which the bag is biased are equally normal, it will be compatible with what you believe after observing the first sequence that you are going to observe the longer sequence. So that subsequent discovery is compatible with your prior beliefs, and occasions a reversal in your beliefs about the bag.

This example makes salient an important fact about probability structures: what they predict is highly sensitive to the choice of Q. An urgent question in assessing the account is then how to think about the status of Q. Addressing this issue head on is something we must defer to future work. But the present project does much to address it indirectly, by seeing what predictions about belief correspond to different choices of Q.[30] Together with judgments about what people believe (or about what 'belief'-sentences are true in what contexts, if the relevant question Q can vary depending on the context – which is our preferred view) these predictions yield nontrivial constraints on any account of the questions relative to which belief is sensitive.[31] We should also note that the idea that belief is question sensitive is not an idea unique to us; it is a feature of most of the alternative probabilistic theories discussed in the next section, and has other precedents in the philosophy of mind and language.[32]

Moreover, we are not suggesting that the question *is the bag red-biased* is the only natural question when thinking about **Bias Detection**. This is because more fine-grained questions are needed to allow the agent to have inductive beliefs about their future observations; for example, that their first twenty draws won't all be black. But we don't think that there are any very natural questions that respect ORTHOGONALITY in this case. In particular, maximally fine-grained questions are unattractive. For when your evidence entails how many draws you will make from the bag, a maximally fine-grained question threatens to make all states in which the bag is unbiased equally normal, preventing any nontrivial beliefs about what red-black sequence will be observed.[33]

---

[30] For a worked example, see Goodman and Salow (2021, §5).

[31] On one version of this view, failures of $\Diamond R$ might be 'elusive', in the sense of Lewis (1996): focusing on what the agent learns might tend to shift the context to one in which their discovery doesn't violate $\Diamond R$, by shifting the question to one congruent with their discovery.

[32] See, e.g., Yalcin (2018), Hoek (forthcoming), Holguín (2022).

[33] While those with skeptical leanings about induction in these cases might embrace this conclusion, fine-graining is not a general recipe for skepticism, and often predicts counterin-

To illustrate how failures of ORTHOGONALITY can generate failures of $\Pi-$, consider the following case:

Celebrity Hike
101 celebrities go on a hike in Runyon Canyon. A paparazzo shadowing them notices a hiking pole on the trail. On inspection, he notices some fingerprints. He knows that Michael Jackson and Beyoncé were on the hike, and that he always hikes wearing one glove on his right hand and she always hikes wearing one glove on her left hand. After inspecting the pole further, the paparazzo discovers whether the fingerprints were made by a left or right hand.

We can model the case using 200 equally probable states compatible with the paparazzo's evidence after he notices the fingerprints: one for each hand that might have made them (Michael's left hand, Beyoncé's right hand, and either hand of any of the other celebrities on the hike). Let Q be *who dropped the pole*, p be that the pole was carried in a left hand, and q be that someone other than Michael or Beyoncé dropped the pole. Whether the paparazzo discovers p or its negation, he won't then believe q, since all answers to Q will be equally likely on his evidence, and he will have no inductive beliefs. But for $t \leq .99$, he will still believe q beforehand, generating a counterexample to $\Pi-$.

9 Other probabilistic approaches

In this section (which can be skipped without loss of continuity) we contrast the account developed so far with other probabilistic accounts of belief dynamics in the literature. We start with the well-known Lockean theory mentioned at the outset, which cannot be accommodated within the normality framework since it recommends having inconsistent beliefs that are not closed under entailment. We then discuss theories based on Leitgeb (2017), Lin and Kelly (2012), and Levi (1967), explaining how these can be viewed as alternative proposals for generating normality structures from probability structures.

tuitive patterns of belief. For example, if our evidence doesn't settle how many times you will draw from the bag, then it tends to predict that our inductive beliefs to the effect that you won't make *more* than a certain number of draws are much stronger (as measured by their evidential probability) than our inductive beliefs to the effect that you won't make *fewer* than a certain number of draws – since the more draws you make the more fine-grained (and hence less probable) the states involved must be in order to take a stand on one's future discoveries.

These contrasting predictions about belief, depending on whether your evidence settles the number of draws, is reminiscent of the sensitivity of classical significance tests to stopping rules. Partition-sensitivity also arises in classical statistics in terms of the choice of test statistics; see Kotzen (2022) for a helpful survey. In general, partition-sensitivity is hard to avoid in any theory that aims to generate a qualitative summary of a probability distribution.

9.1 Lockeanism

Let us start by distinguishing two principles:

> THRESHOLD: You believe that p only if the probability of p, given your evidence, is at least t.

> LOCKEANISM: You believe that p if and only if the probability of p, given your evidence, is at least t.

The view we have developed embraces THRESHOLD but rejects LOCKEANISM: high probability is necessary but not sufficient for belief.

The synchronic predictions of LOCKEANISM are sensitive to the choice of t. As is well-known, if $t < 1$, the view allows you to believe two things without believing their conjunction, preventing your beliefs from being characterized by a single set of states compatible with what you believe. Even if $t = 1$, one's beliefs can still be inconsistent. For example, the probability on Bjorn's evidence that he weighs exactly x pounds is arguably zero for each $x \in \mathbb{R}$. So Lockeans cannot model belief as truth in all doxastically accessible possibilities.

In exploring the dynamic predictions of LOCKEANISM, we will assume that probabilities evolve by conditionalization on what one learns (and set aside cases where what one learns had probability 0). If $t = 1$, LOCKEANISM then vindicates all the principles we have discussed, including principles such as $\Diamond-$, which we have argued against at length. If $t < 1$, things are more interesting.[34]

Proposition 13. LOCKEANISM *validates* $\Pi-$, $\Pi+$, *and* FRONTLOADING; *and* $\Box R$ *holds whenever* $t > \frac{\sqrt{5}-1}{2} \approx .62$.

Proposition 14. $\Diamond-$, $\Diamond R$, $\Box+$, $\Box-$ *can all fail given* LOCKEANISM *for* $t < 1$.

The belief revision theory of LOCKEANISM treats principles of belief preservation and principles of belief acquisition symmetrically: $\Pi-$ and $\Pi+$ stand together, while $\Box-$ and $\Box+$ fall together. As noted at the end of §7 this constitutes a difference from normality structures derived from simple probability structures, which are friendlier to $-$ principles than to $+$ principles.

We reject LOCKEANISM, but not because of Propositions 13 and 14. Our (not at all original) concern with LOCKEANISM is rather that its static predictions are implausible. According to the view, Bjorn should believe that he does not weigh x pounds for every x, no matter what he sees on his scales; when he sees a single measurement of y, he should (for suitably small d) believe that he weighs either less than $y - d$ or more than $y + d$. Similarly, in a version of **Bias Detection** where you are likely to draw from the bag a large number of times, it says you should believe of every possible outcome that it will not obtain. In **Flipping for Heads**, LOCKEANISM implausibly says that you should believe that the coin will be flipped at most n times and believe that it won't be flipped exactly n

---

[34]Hawthorne (1996) and Shear and Fitelson (2019) systematically study nonmonotonic consequence and belief revision, respectively, in a Lockean framework; Appendices D.1-D.2 explain how to translate from these frameworks into ours.

times, but not believe that it will be flipped at most $n - 1$ times. LOCKEANISM also severs the link between rational belief and potential knowledge, discussed in §3.1: the most normal possibilities compatible with one's evidence cannot be known not to obtain, but they may have a low enough probability individually that LOCKEANISM says to believe that they don't obtain.

The main advantage of LOCKEANISM is that, by denying that we believe the conjunction of everything we believe, it can reconcile THRESHOLD with there being many propositions, about independent questions, that we believe and whose probability is closer to the threshold t than it is to 1.[35] While this can't happen according to our view, we can accept the parallel meta-linguistic claim that there are many propositions, about independent questions, that we can be truly said to "believe" and whose probability is closer to the threshold t than it is to 1. As alluded to at the end of the last section, and discussed at greater length in Goodman and Salow (2021, 2023, §§5), the idea is that "belief"-reports are context-sensitive, and making such reports can change the context to one associated with questions more favorable to the truth of those reports.

9.2   Stability Theories

This section explores the *stability theory* of belief developed and defended by Leitgeb (2017), as well as a simpler proposal inspired by it. The simpler proposal can be modeled using probability structures, while Leitgeb's view requires enriching those structures with evidence-dependent thresholds.

We begin with the simpler proposal. Say that p is *topical* just in case $p = \bigcup X$ for some non-empty $X \subseteq Q$, and that a probability structure is *topical* just in case $\mathcal{E} = \{p : p \text{ is topical}\}$. Following Leitgeb, say that p is *r-stable for Pr* just in case $Pr(p|q) > r$ for all topical q consistent with p; and say that p is *weakly r-stable* just in case $Pr(p|q) \geq r$ for all topical q consistent with p. Now consider the following constraints on normality structures determined by probability structures:

> .5-STABILITY
> For all $E \in \mathcal{E}$, $B(E)$ is .5-stable for Pr_E.

> t-STABILITY
> For all $E \in \mathcal{E}$, $B(E)$ is weakly t-stable for Pr_E.

> WEAK LOCKEANISM
> For all $E \in \mathcal{E}$, there is some r such that, for all topical p, $B(E) \subseteq p$ if and only if $Pr_E(p) \geq r$.

Proposition 15. *.5-*STABILITY *and* WEAK LOCKEANISM *are equivalent constraints on normality structures determined by probability structures.*

Proposition 16. $\Diamond-$ *and t-*STABILITY *are equivalent constraints on normality structures determined by topical probability structures.*

---

[35]There are ways to generalize probability structures to achieve this effect while avoiding the unattractive predictions of LOCKEANISM mentioned above; see Goodman (2024, appendix B).

Proposition 15 is the transposition to the present setting of one of Leitgeb's core results. Together with proposition 16 (and the fact that being weakly r-stable entails being r'-stable whenever $r > r'$), it means that, in topical structures with $t > .5$, $\Diamond-$ (a dynamical constraint on belief) is sufficient for WEAK LOCKEANISM (a synchronic constraint linking belief and probability). This result parallels Leitgeb's informal motivation for his own theory, namely a desire to accept a stability requirement on belief and a version of Lockeanism in a setting where belief is closed under entailment.

While WEAK LOCKEANISM has some intuitive appeal (saying, in effect, that one believes any topical claim at least as probable as any other one believes) we do not regard the above results as reason to reconsider our case against $\Diamond-$. If anything, we think that the connection to stability casts the unpalatable skeptical implications of $\Diamond-$ in especially sharp relief: for example, as Kelly and Lin (2021) point out, in **Flipping for Heads**, the requirement that $B(E)$ be .5-stable for Pr_E is tantamount to requiring that $t = 1$.[36] In our view, any intuitive pull there is to .5-STABILITY and WEAK LOCKEANISM is outweighed by anti-skeptical considerations in cases like **Flipping for Heads** and (especially) **Bias Detection**.

Interestingly, neither t-STABILITY nor .5-STABILITY entails $\Box+$.[37] This shows that there are principled (if not, in our view, plausible) bases for maintaining $\Diamond-$ without accepting the full AGM orthodoxy.

We now turn to Leitgeb's (2017) stability theory. Let a *flexible probability structure* be just like a probability structure, except that the threshold is a function of the agent's evidence; and let $\gg_E$ and $B(E)$ be determined just as in probability structures, except with t_E in place of t. It turns out that Leitgeb's way of modeling belief is equivalent to using normality structures determined by flexible probability structures that satisfy two further constraints.

The first constraint is .5-STABILITY (equivalently, WEAK LOCKEANISM). The second constraint is, in effect, that AGM holds: structures must be topical and the threshold must vary with one's evidence in such a way that $B(E \cap p) = B(E) \cap p$ for all topical p compatible with $B(E)$.[38] Note that the first constraint alone does not entail $\Diamond-$,[39] much less AGM, while the second constraint alone is compatible with one's beliefs having arbitrarily low positive probability.

Recasting Leitgeb's theory in terms of flexible probability structures reveals something important about its structure: .5-STABILITY and WEAK LOCKEANISM,

---

[36] Rott (2017) and Douven and Rott (2018) raise similar concerns.

[37] Consider a simple probability structure with $S = \{a, b, c\}, \mathcal{E} = \mathcal{P}(S) \setminus \{\emptyset\}, Pr(\{a\}) = .9, Pr(\{b\}) = .099, Pr(\{c\}) = .001, t = .9001$. $\Box+$ fails, since $B(S) = \{a, b\} \not\subseteq \{a\} = B(\{a, b\})$.

[38] Here is a recipe for constructing flexible probability structures that obey this constraint. Start by choosing any threshold whatsoever for t_S, which determines $B(S)$. Then, for every other E consistent with $B(S)$, let $t_E = Pr_E(B(S))$. Next, choose any threshold whatsoever for $t_{S \setminus B(S)}$, corresponding to the weakest body of evidence not yet assigned a threshold. As before, this determines $B(S \setminus B(S))$ and thereby fixes t_E for all other E consistent with $B(S \setminus B(S))$. Repeat this procedure until every topical E is assigned a threshold.

[39] Consider a simple probability structure with $S = \{a, b, c\}, \mathcal{E} = \mathcal{P}(S) \setminus \{\emptyset\}, Pr(\{a\}) = .5, Pr(\{b\}) = .3, Pr(\{c\}) = .2$, and $t_E = .8$ for all $E \in \mathcal{E}$. $B(E)$ is .5-stable for Pr_E for all $E \in \mathcal{E}$. $\Diamond-$ fails: $B(\{a, c\}) = \{a, c\} \not\subseteq \{a, b\} = B(S)$.

on the one hand, and AGM, on the other, are really two independent commitments. (For this reason, the word 'stability' strikes us as a somewhat unfortunate choice for what is essentially a synchronic constraint, given its diachronic connotations.) This isn't intended as an objection: our reason for rejecting Leitgeb's theory is not any worry about the disunity in how WEAK LOCKEANISM and AGM are motivated, but rather the fact that .5-STABILITY and $\lozenge-$ each have similar implausible consequences.

9.3   The Tracking Theory

Lin and Kelly (2012) defend what they call the 'tracking theory' of belief. It is so called because it allows for belief dynamics and probability dynamics to march in step in a sense they make precise.[40] This theory can be seen as an alternative way of determining normality structures from probability structures, with the parameter t playing a rather different role.

Definition 9.1. $\langle S, \mathcal{E}, W, \succeq, \gg \rangle$ is *LK-determined by* the probability structure $\langle S, \mathcal{E}, W, Q, Pr, t \rangle$ if and only if

- $\succeq_E$ is as in Definition 8.2;

- $s \gg_E s'$ iff $\frac{Pr([s]_Q|E)}{Pr([s']_Q|E)} > \frac{1}{t}$;

- situations involving different evidence are not related by $\succeq$ or $\gg$.

LK-determination generates normality structures from probability structures. The intuition behind them is that a situation is sufficiently more normal than another if the answer to Q that is true in that situation is sufficiently more likely than the answer to Q that is true in the other situation. In many cases – such as **Flipping for Heads** (and, when generalized to probability densities, **Bjorn**) – LK-determined normality structures will generate similar predictions about belief to those of our preferred models (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, $\gg$ is LK-determined locally, in the sense that whether $s \gg_E s'$ depends only on the probabilities (given E) of $[s]_Q$ and $[s']_Q$. As a result:

Proposition 17. *Normality structures LK-determined by probability structures satisfying* ORTHOGONALITY *satisfy* STATISM *in addition to* COMPARABILITY.

So given ORTHOGONALITY (which is arguably a presupposition of Lin and Kelly's models), the tracking theory is an instance of the kind of view mentioned in §4, on which $\lozenge-$ fails, but $\square+$, $\square-$, $\Pi+$, $\Pi-$, and $\lozenge R$ all hold, owing to the fact that COLLAPSE can fail but COMPARABILITY and STATISM cannot. As discussed

---

[40]Kelly and Lin (2021) develop a related theory, designed to accommodate what they call 'symmetric Gettier cases': cases where a and b are equally plausible and each is more plausible than the only other possibility c in a way that licenses you to believe that one of a and b obtains; but where the difference in plausibility between a and c individually isn't big enough for you to believe that a obtains after learning b does not (which would involve violations of STATISM).

earlier, this is a fairly minimal and seemingly principled departure from AGM, which one might reasonably take to tell in favor of the theory.

The major shortcoming of the tracking theory, in our view, is that it fails to entail THRESHOLD. Consider a case like **Computer Number**, in which one state has very low probability (10^{-5}) but every other state has extremely low probability ($< 10^{-10}$), and let $Q = \{\{s\} : s \in S\}$ be the question *which state obtains*. Even for extremely low values of t ($> 10^{-5}$) – values whose implications in cases like **Flipping for Heads** are borderline skeptical – the tracking theory recommends believing that the computer is broken, even though this has probability 10^{-5} on your evidence.

Similar problems arise in **Bias Detection**. Suppose you have decided to draw from the bag a large number of times. If we consider the fine-grained question, which settles both the bias of the bag and the exact sequence of red/black draws, the tracking theory says to believe that the bag is biased and that almost all the draws will be red, since the most likely outcome is that the bag is biased and all draws are red, and this is sufficiently more likely than any outcome on which the bag is unbiased or on which it is biased but a significant number of draws are black. But given your evidence it is extremely unlikely that the bag is biased and almost all the draws will be red. This should not be something you believe.

One might defend the tracking theory against these counterexamples 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 of the principle just discussed. However, moving to coarser-grained questions requires rejecting ORTHOGONALITY, which as discussed earlier fails in **Bias Detection** for natural coarser-grained questions such as *is the bag biased* and *is the bag biased and what proportion of draws will be red*. And without ORTHOGONALITY, the tracking theory doesn't satisfy STATISM (or WEAK STATISM) any more than our way of determining normality structures from probability structures does, significantly weakening the resulting logic of belief revision:

Proposition 18. $\Box+$ *and* $\Box-$ *are valid on the class of normality structures LK-determined by probability structures; but* $\Pi+$, $\Pi-$, *and* $\Diamond R$ *can all fail.*

Without ORTHOGONALITY, $\Box+$ is the only principle valid on the class of normality structures LK-determined by probability structures but not on the class of structures determined (in the sense of Definition 8.2) by probability structures. Moreover, the tracking theory now makes some truly bizarre predictions. Consider the following example, which Hacking (1967) introduced to make a parallel objection to Levi (1967):

> **Drawing a Card**
> Before you are 65 decks of cards. 52 of these (one for each combination of number and suit) are trick decks, containing 52 copies of the same card. The other 13 decks are fair. You select a deck at random, shuffle it, and draw a card.

Let $Q =$ *which of the 53 possible deck-types did you select* and $t > .25$. According to the tracking theory, you initially believe that you selected a fair deck, but after drawing a card you believe that you selected the relevant trick deck. So we have a failure of the following principle:

ΠR If you believe q, then for any set of mutually exclusive and jointly exhaustive discoverable events, it is possible to discover one of them without coming to believe not-q.

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

By contrast, as long as belief requires probability over a threshold greater than .5, this principle cannot fail.[41]

Overall, then, we see few advantages for the tracking theory over ours. The fact that it uses a local condition to define $\gg$ means that it can vindicate STATISM given ORTHOGONALITY, yielding a strong logic of belief revision. However, this locality prevents the theory from entailing THRESHOLD, which is a global constraint on the total probability of the set of doxastic possibilities. Moreover, to make reasonable predictions in cases like **Bias Detection**, both theories need to appeal to coarse-grained questions that conflict with ORTHOGONALITY. Both frameworks then predict violations of even WEAK STATISM, yielding different but significantly weaker logics of belief revision in both cases.[42]

9.4   Levi's theory

Levi (1967) developed an influential theory based on the idea that one's beliefs should offer the best trade-off between being informative and being probable. Again, we can interpret his theory as an alternative proposal for determining normality structures from probability structures:

Definition 9.2. $\langle S, \mathcal{E}, W, \succeq, \gg \rangle$ is *Levi-determined by* the probability structure $\langle S, \mathcal{E}, W, Q, Pr, t \rangle$ if and only if Q is finite and

- $\succeq_E$ is as in Definition 8.2;

- $s \gg_E s'$ iff $Pr([s]_Q | E) \geq \frac{t}{|\{q \in Q : q \cap E \neq \emptyset\}|} > Pr([s']_Q | E)$;

- situations involving different evidence are not related by $\succeq$ or $\gg$.

Levi-determined structures satisfy the conditions for being normality structures. Intuitively, here is what these structures say one should believe. Levi glosses t

---

[41] Failures of ΠR 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 kind of deck you selected is going to change no matter what card you draw. For more on this notion of belief see Holguín (2022); cf. the account of good guesses in Dorst and Mandelkern (2023).

[42] Goldstein and Hawthorne (2022) adopt a hybrid account of $\gg$ that combines the global probability comparisons of definition 6.2 with the local ones of definition 9.1: $\langle s, E \rangle \gg \langle s', E \rangle$ iff $Pr(\{s'' : s' \not\succeq_E s''\} | \{s'' : s \succeq_E s''\}) \geq t$ and $\frac{Pr([s]_Q)}{Pr([s']_Q)} > \frac{1}{t'}$. We are ambivalent about this account: it has some attractions (Goodman and Salow, 2021, footnote 10), but it forsakes the natural direct characterization of belief in Proposition 9. It makes little difference to the logic of belief revision, as versions of Propositions 10-12 continue to hold on the hybrid account.

as a measure of 'boldness'. When $t = 1$, agents are maximally bold and their beliefs rule out all and only the cells of Q with below-average probability. More generally, agents' beliefs rule out exactly the answers to Q with probability less than t times the average probability of the answers to Q that are compatible with their evidence. So defined, $\gg$ is extremely coarse grained: relative to a possible body of evidence, it corresponds to a normal/abnormal dichotomy of states compatible with that evidence. Nevertheless, for suitable choices of Q and t, Levi-determined normality structures make similar predictions about belief to our preferred models in cases like **Flipping for Heads**.

Like our own probabilistic account, Levi's account makes $\gg$ sensitive to global probabilistic features of the states compatible with one's evidence (in his case, how many answers to Q they are compatible with); so, as with our account, STATISM can fail even assuming ORTHOGONALITY. The resulting situation is largely parallel to our Propositions 8 and 10:

Proposition 19. *Normality structures Levi-determined by probability structures satisfying* ORTHOGONALITY *satisfy* COMPARABILITY *and* WEAK STATISM, *and validate* $\square-$, $\Pi-$, *and* $\lozenge R$.

Proposition 20. $\square+$ *(and hence* STATISM*) can fail in such structures.*

However, the effect of relaxing ORTHOGONALITY is more dramatic on Levi's account than on ours:

Proposition 21. $\lozenge R$, $\square-$, *and* ΠR *can all fail in normality structures Levi-determined by probability structures.*

Our reasons for preferring our account to Levi's – largely anticipated by Hacking (1967) – are virtually the same as our reasons for preferring our account to the tracking theory. Like the tracking theory, Levi's account fails to entail THRESHOLD; and it makes essentially the same predictions about the examples described in the previous subsection. If we impose ORTHOGONALITY, then it predicts believing extremely improbable events in cases like **Bias Detection**. And if we reject ORTHOGONALITY, the resulting belief dynamics are even less constrained than on our account. Given the plausibility of THRESHOLD, our theory thus seems preferable.

10 *De se* questions, and how $\square-$ can fail

In this section we discuss a natural generalization of probability structures. These more general structures determine normality structures in which $\square-$ can fail. We then show how these structures can be used to model cases in which there are independent grounds to think that $\square-$ really does fail.

In probability structures, the partition Q of S corresponds to a question about which state of the world obtains. We can generalize these models by instead appealing to a partition $\mathcal{Q}$ of W, corresponding to a question about

32

what situation one is in. Let us begin by formalizing this idea before applying it to an illustrative example.

Definition 10.1. A *generalized probability structure* is a tuple $\langle S, \mathcal{E}, W, \mathcal{Q}, Pr, t \rangle$ where $\mathcal{Q}$ is a partition of W, and S, $\mathcal{E}$, W, Pr, and t are defined as before.

Strictly speaking, probability structures are not generalized probability structures. But since every partition Q of S determines a unique partition $\mathcal{Q}$ of W (where $[\langle s, E \rangle]_{\mathcal{Q}} := \{\langle s', E' \rangle : s' \in [s]_{Q}\}$), we harmlessly speak as if they are.

To generate normality structures from generalized probability structures, we modify Definition 8.2 by replacing Q in the characterization of $\succeq_E$ with the partition $\mathcal{Q}_E$ of E induced by $\mathcal{Q}$ (where $[s]_{\mathcal{Q}_E} = \{s' : \langle s', E \rangle \in [\langle s, E \rangle]_{\mathcal{Q}}\}$):

Definition 10.2. $\langle S, \mathcal{E}, W, \succeq, \gg \rangle$ is *determined by* the generalized probability structure $\langle S, \mathcal{E}, \mathcal{Q}, W, Pr, t \rangle$ if and only if:

- $s \succeq_E s'$ iff $Pr([s]_{\mathcal{Q}_E} | E) \geq Pr([s']_{\mathcal{Q}_E} | E)$;

- $\gg_E$ is as in Definition 6.2;

- situations involving different evidence are not related by $\succeq$ or $\gg$.

This is a generalization of Definition 8.2 because it yields the same normality relations as the old definition in the special case where $\mathcal{Q}$ is determined by a partition of states Q as described above.

Generalized probability structures offer greater flexibility than probability structures by, in effect, making the relevant question about the state of the world be a function of one's evidence. This in turn has consequences for the logic of belief revision:

Proposition 22. $\square-$ *can fail in normality structures determined by generalized probability structures.* ΠR *is valid provided $t > .5$ and $\square R$ is valid provided $t > \frac{\sqrt{5}-1}{2} \approx .62$.*

To see why this increased flexibility is desirable, consider the following cases from Goodman and Salow (2023a, §8):

> **Flipping for All Heads**
> A coin flipper will simultaneously flip 100 fair coins until they all simultaneously land heads. Then he will flip no more.

> **Decay**
> A radioactive atom is created; eventually, it will decay. The average time for an atom of this isotope to decay is one year.

Let's consider what one should believe about how long it will take before all of the coins land heads together in **Flipping for All Heads**. (Analogous claims seem plausible concerning your beliefs about what day the atom will decay in **Decay**.) A natural thought is that your beliefs should be two-sided: you believe that it will happen eventually (i.e. before the yth flip, for some large y) but not

soon (i.e. not before the xth flip, for some $1 < x < y$). Moreover, the shape of these beliefs remains the same as you observe that it hasn't happened yet: if you see that the coins don't all land heads on any of the first n flips, you should afterwards believe that they won't all land heads before the $(n + x)^{th}$ flip but will all land heads before the $(n + y)^{th}$ flip.

If these natural claims about **Flipping for All Heads** are correct, then we have a counterexample to $\Box-$. For you initially believe that the coins won't all land heads before the yth flip; you stop believing this when you see that they don't all land heads on the first flip; and yet you initially believed that they wouldn't all land heads on the first flip.

We can model this failure of $\Box-$ using a generalized probability structure:

$$S = \{s_1, s_2, s_3 \ldots\}.$$

$$\mathcal{E} = \big\{E_i := \{s_i, s_{i+1}, s_{i+2} \ldots\} : i \geq 1\big\}.$$

$$\mathcal{Q} = \Big\{\{\langle s_j, E_i\rangle : j{-}i < x\}, \{\langle s_j, E_i\rangle : x \leq j{-}i \leq y\}, \{\langle s_j, E_i\rangle : y < j{-}i\}\Big\}.$$

$$Pr(\{s_i\}) = (1 - \tfrac{1}{2^{100}})^{i-1}(\tfrac{1}{2^{100}}).$$

In this model, s_i is the state in which the coins all land heads together for the first time on flip i; the possible bodies of evidence settle, at each point, that the coins haven't all landed heads together before then; the (three-answer) question is *when will the coins all land heads together: extremely soon, an extremely long time from now, or in between*; and the probability distribution matches the objective chances before any flips take place. Provided that x is relatively small and y is very big (with the exact values depending on the threshold t), the subject will believe at each time (before the thrilling experiment comes to an end) that the coins will all land heads together neither extremely soon nor extremely long from then. As time passes, this has different implications for how long from the start of the experiment it will be before the coins all land heads together, which is what leads to counterexamples to $\Box-$.

The above model is crude, but the broad shape of its predictions survive further refinements. For example, a more realistic question might be a little more fine-grained, perhaps something along the lines of: *will it happen on one of the next 1-10 flips, or on one of the next 11-100 flips, or on one of the next 101-1000 flips, etc.* This will yield qualitatively similar predictions. In Goodman and Salow (2021, Appendix C), we show that there is a mathematically natural way to make qualitatively similar predictions in **Decay** using the extremely fine-grained continuum-valued question *how much longer until it decays*, provided that we use the right probability density function.

We call $\mathcal{Q}$ a *de se* question because our situations play a role similar to the 'centered worlds' of Lewis (1979), so partitions of W are naturally thought of as questions that take a stand not only on how the world is but also on where in the world you are. We make these connections more precise in Appendix B. There we introduce a natural generalization of normality structures to a multi-agent setting, which requires explicitly identifying situations with Lewisian centered

worlds (i.e. state/time/agent-triples) with the upshot that no extra machinery is needed to model self-locating evidence and inductive belief (i.e., evidence and beliefs concerning not just the state of the world but one's place in it).

11   Conclusion

We hope to have illustrated the fruitfulness of normality structures for theorizing about the dynamics of rational belief. They offer natural models of belief revision, and interface with recent work in traditional epistemology, epistemic logic, and Bayesian epistemology.

We have also applied the framework to a range of examples. Although few principles of belief revision have emerged unscathed, three interesting levels of idealization have emerged. These are (i) the simple probability structures of §6 (or, equivalently, probability structures satisfying ORTHOGONALITY), (ii) the more general class of probability structures motivated in §8, and (iii) the still more general class of structures involving *de se* questions described in §10. $\Box R$ is valid in all three model classes given the natural assumption that rational belief requires probability at least $\frac{\sqrt{5}-1}{2}$ ($\approx .62$). But $\Box-$ can fail in the last class of structures, $\Diamond R$ can fail in the the second and third classes of structures, and $\Diamond-$ and $\Box+$ can fail in all three classes of structures.

In our view, there is no natural level of idealization for which either $\Diamond-$ or $\Box+$ holds. This is less surprising in the case of $\Diamond-$, where there are purported counterexamples in the literature and, in our view, a decisive counterexample in **Bias Detection**. By contrast, our case against $\Box+$ (and FRONTLOADING) is more theoretical. Its failure falls out naturally from probabilistic accounts of normality, and also from the compelling thought that, typically, strengthening your evidence makes it reasonable to form new inductive beliefs. In our view, the intuitive appeal of $\Box+$ is an artifact of operating in frameworks that obscure the distinction between evidence and inductive belief. One of the most important features of normality structures is keeping that distinction front and center.

	$\Diamond-$	$\Diamond R$	$\Box+$	$\Box-$	$\Box R$	II+	II−	IIR	COMP.	COLL.	STAT.	W. STAT.	THRE.
AGM	✓	✓	✓	✓	✓	✓	✓	✓					
STATISM, COLLAPSE, and COMPARABILITY	✓	✓	✓	✓	✓	✓	✓	✓					
STATISM and COMPARABILITY	✗	✓	✓	✓	✓	✓	✓	✓					
COMPARABILITY	✗	✓	✓	✓	✓	✓	✗	✓					
STATISM	✗	✓	✗	✓	✓	✗	✗	✓					
WEAK STATISM	✗	✓	✗	✗	✓	✗	✗	✓					
Probability structures and ORTHOGONALITY	✗	✓	✗	✓	✓	✗	✓	✓	✓	✗	✗	✓	✓
Probability structures	✗	✗	✗	✓	✓	✗	✗	✓**	✓	✗	✗	✗	✓
Generalized probability structures	✗	✗	✗	✗	✓*	✗	✗	✓**	✓	✗	✗	✗	✓
LOCKEANISM	✗	✗	✗	✗	✓*	✓	✓	✓	-	-	-	-	✓
t-STABILITY and ORTHOGONALITY	✓	✓	✗	✓	✓	✗	✓	✓	✓	✗	✗	✓	✓
Tracking Theory and ORTHOGONALITY	✗	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✗
Tracking Theory	✗	✗	✓	✓	✓	✗	✗	✗	✓	✗	✗	✗	✗
Levi's Theory and ORTHOGONALITY	✗	✓	✗	✓	✓	✗	✓	✓	✓	✗	✗	✓	✗
Levi's Theory	✗	✗	✗	✗	✓	✗	✗	✗	✓	✗	✗	✗	✗

Summary of Main Results: ✓* means that the theory vindicates the principle if the probability threshold $t > \frac{\sqrt{5}-1}{2} \approx .62$. ✓** means that the theory vindicates the principle if the probability threshold $t > .5$. Note that Lin and Kelly (2012) assume ORTHOGONALITY in formulating their Tracking Theory, but Levi does not assume ORTHOGONALITY in formulating his theory. Proofs in Appendix E.

A   Learning what you've learned

An important moral of thought experiments like the Monty Hall problem is that, when we learn something about the world, we also learn that we have learned it. Accommodating this insight using normality structures turns out to correspond to a non-trivial constraint on the possible bodies of evidence, with substantive implications for the logic of belief revision.

To illustrate the issue, consider the following example:

> **Babysitter**
> Alex and Carole say that they probably can't both come to your party: they'll both come if they can find a babysitter, but otherwise they will flip a coin to decide which of them will go. All the other guests have arrived, and the doorbell rings: you are about to look through the peep hole and you know you're about to see either of Alex or Carole (whoever rang the bell).

It is tempting to model this case using a simple probability structure with three states: b, in which they find sitter and both come, a in which they don't find a sitter and only Alex comes, and c in which they don't find a sitter and only Carole comes. The two potential discoveries are then $\{a, b\}$, which is what you learn if you see Alex, and $\{c, b\}$, which is what you learn if you see Carole. Suppose that, before answering the door, b has probability .1 and a and c each have probability .45. So b has probability $\frac{2}{11}$ conditional on either of the two potential discoveries: whatever you learn, the probability that Alex and Carole found a sitter will increase. For $t = .9$, this would mean that you will initially believe that they won't find a sitter, but you will give up this belief whoever you see through the peephole.[43]

This is clearly the wrong result: assuming that Alex and Carole are equally likely to ring the doorbell if they both come, the probability that they found a sitter shouldn't change when you see one of them through the peephole, and certainly shouldn't go up no matter what, since you were certain you were going to see one of them.

The problem is that the model individuates states too coarsely, in a way that forces it to distort the content of your evidence. When you look through the peephole, you don't learn only that Alex came or that Carole came: you also learn that they rang the doorbell. Since whoever rang the doorbell was certain to do so if they came alone, but only .5 likely to do so if they came together, this additional evidence is relevant to whether they came together, and taking it into account leaves the probability that they came together unchanged.

We can capture this fact in a simple probability structure by individuating states more finely. We replace the single state b with two states b_a and b_c, which differ in who you see through the peephole and which both have probability .05. The possible discoveries are then $\{a, b_a\}$ and $\{c, b_c\}$. The probability that they

---

[43]Smith (2018b) uses an example similar to **Babysitter** to argue against Lockeanism on the grounds that such belief dynamics are intolerable.

both came – i.e., of $\{b_a, b_c\}$ – is then unchanged by conditioning on either of these two discoveries.[44]

Reflection on this case suggests a general constraint on normality structures. Typically, when you make a discovery, you thereby learn that you have made that discovery. Accommodating such learning in normality structures requires that states be sufficiently fine-grained to settle what discoveries you will make. This is tantamount to the requirement that, for any situation $\langle s, E \rangle$, s settles that at some point your total evidence about the state of the world is E. Equivalently, states take a stand on your *evidential trajectory*: there is a function τ that maps each state s to the set of possible bodies of evidence that are ever your total body of evidence when s obtains. Since E is your evidence in $\langle s, E \rangle$, $E \in \tau(s)$ for all $s \in E \in \mathcal{E}$. And setting aside the possibility of memory loss or otherwise losing evidence (a reasonable idealization in the cases we are considering here), all evidential changes are discoveries and your evidence only ever increases. So your evidential trajectory is linearly ordered by logical strength: if $E, E' \in \tau(s)$, then either $E \subseteq E'$ or $E' \subseteq E$.

The existence of such a function τ mapping states to evidential trajectories is equivalent to the following constraint on possible bodies of evidence:[45]

NESTEDNESS: If $E, E' \in \mathcal{E}$ and $E \cap E' \neq \emptyset$, then $E \subseteq E'$ or $E' \subseteq E$.

The problem with the original three-state model of **Babysitter** is that it failed to respect this constraint.[46]

Imposing NESTEDNESS has further implications for the logic of belief revision. Consider the following strengthening of $\Pi-$:

C$-$ If $\mathcal{C} \subseteq \mathcal{E}$ is finite with $\bigcup \mathcal{C} = E$, then $B(E \cap p) \subseteq B(E)$ for some $p \in \mathcal{C}$.

---

[44]**Babysitter** is closely related to the Monty Hall problem, assuming Monty (i) always opens a door you didn't select, (ii) always opens a losing door, and (iii) chooses randomly which door to open when you have selected the winning door. Suppose that, on this occasion, you select Door 2 and Monty then reveals Door 1 to be a losing door. If your evidence is exhausted by the fact that Door 1 is a losing door, this should cause you to become .5 confident that Door 2 is the winning door, generating the incorrect verdict that there is no reason to switch. If, instead, we think of your evidence as also containing the fact that Monty opened Door 1 – which had probability 1 conditional on Door 3 being the winning door but only probability .5 conditional on Door 2 being the winning door – this should leave your confidence that Door 2 is the winning door unchanged at $\frac{1}{3}$, generating the correct verdict that you should switch.

[45]This principle about possible bodies of evidence should be distinguished from the principle defended by Dorst (2020) that evidential accessibility is (shift-)nested: if $u, v \in R_E(w)$ and $R_E(u) \cap R_E(v) \neq \emptyset$, then $R_E(u) \subseteq R_E(v)$ or $R_E(v) \subseteq R_E(u)$. This holds trivially in normality structures because in such structures evidence is transparent (although we agree with Williamson (2019) and Das (2023) that the principle is implausible once that idealization is relaxed). The idealization underpinning NESTEDNESS is stronger, since it can fail in cases of memory loss where the transparency of evidence is still a reasonable idealization.

[46]The same is true of our model of **Flipping for Heads** in §7, which accommodated the possibility of being informed whether the coin landed heads in the first n flips by simply adding $\{s_1, \ldots, s_n\}$ to $\mathcal{E}$, thereby violating NESTEDNESS. The resulting model is therefore unrealistic. As in **Babysitter**, the solution is to divide states according to whether you are going to watch the coin as it is flipped or wait to be informed later whether it has landed heads in the first n flips. Moving to this more realistic model doesn't disrupt any of our earlier arguments.

We saw that this principle failed in the naïve three-state model of **Babysitter**. But it is equivalent to $\Pi-$ given NESTEDNESS, and so holds in the modified four-state structure (since $\Pi-$ is valid on the class of normality structures determined by simple probability structures).[47]

B   Multi-agent models and self-location

In this appendix we explain how to give a multi-agent generalization of our models. Doing so is not entirely straightforward, for the following reason. As explained in § 2.1, in order to model learning as a decrease in which possibilities are compatible with one's evidence, we need to model these possibilities as states rather than situations. (This is because situations take a stand on what your evidence is, so after getting new evidence a completely new set of situations is compatible with your evidence: ones in which you have different, stronger evidence about the state of the world.) This doesn't create any problems in the single-agent case, because we are only interested in agents' beliefs about the state of the world, and not in their beliefs about their current evidence.

But in the multi-agent case, we are interested in agents' beliefs about agents' evidence: even if we're working at a level of idealization where every agent's evidence is transparent to that agent, different agents' evidence isn't transparent to each other.

A natural and conservative way to handle this issue is to work with the following more general class of structures:

Definition B.1. A *de se normality structure* is a tuple $\langle S, T, A, W, R_E, \succeq, \gg \rangle$ such that:

1. S is a non-empty set (of *states*),

2. T is a non-empty set (of *times*),

3. A is a non-empty set (of *agents*),

4. $W = S \times T \times A$

5. $R_E : W \to \mathcal{P}(W)$ such that

   (a) $w \in R_E(w)$
   (b) If $v \in R_E(w)$, then $R_E(v) = R_E(w)$.

6. $\succeq$ and $\gg$ are as in the definition of normality structures.

The basic notion of evidence in *de se* normality structures is evidence about what situation one is in, represented by R_E. By contrast, in (single agent) normality structures, the basic notion of evidence is evidence about the state,

---

[47]Given NESTEDNESS, we also have that $\Pi+$ entails the following strengthening:

C+   If $\mathcal{C} \subseteq \mathcal{E}$ is finite with $\bigcup \mathcal{C} = E$, then $B(E) \subseteq \bigcup_{p \in C} B(E \cap p)$.

represented by $E \in \mathcal{E}$. We can recover such a notion in *de se* normality structures as $E(w) = \{s : \langle s, t, a \rangle \in R_E(w)$ for some $t \in T$ and $a \in A\}$. We do not assume that one's evidence entails who one is or what time it is. These structures can thus straightforwardly model inductive belief under self-locating uncertainty.

As before, we can restrict our attention to beliefs and discoveries about the state of the world: $B(w) = \{s : \langle s, t, a \rangle \in R_B(w)$ for some $t \in T$ and $a \in A\}$, with R_B defined as in Definition 2.2. A situation $\langle s', t', a' \rangle$ is *the result of learning p in* $\langle s, t, a \rangle$ just in case $s = s'$, $a = a'$, and $E(\langle s', t', a' \rangle) = E(\langle s, t, a \rangle) \cap p$. These notions allow us to interpret principles about belief revision in *de se* normality structures: for example, $\Diamond-$ says that, if $B(w) \cap p \neq \emptyset$ and v is the result of learning p in w, then $B(v) \subseteq B(w)$.

It is straightforward to define an analogue of probability structures for *de se* normality structures, allowing $\succeq$ and $\gg$ to be determined by a prior Pr over S, a partition Q of S, and a threshold $t \in (0, 1]$ as before. However, in defining an analogue of generalized probability structures, there is more flexibility: for example, we might want Pr in addition to Q to be defined over W rather than S (to accommodate non-trivial self-locating prior probabilities) and/or for Pr to be a function of A (in order to capture the idea that different agents have different priors).

C   Non-transparent evidence

One of the major (if contested) morals of recent work on skepticism about the external world is that evidential accessibility isn't symmetric: when we are misperceiving, our evidence is compatible with our perceiving, but when we are perceiving, our evidence is incompatible with things not being as we perceive them to be; cf. Williamson (2000, chapter 8). In this appendix we consider how we might modify normality structures to accommodate this kind of evidential asymmetry.

To theorize about these cases, we introduce the following generalization of normality structures:

Definition C.1. A *generalized normality structure* is a tuple $\langle S, W, R_E, \succeq, \gg \rangle$ such that:

1. $S, \succeq$, and $\gg$ are as in the definition of a normality structure,

2. $W \subseteq \{\langle s, E \rangle : s \in E \subseteq S\}$,

3. $R_E : W \to \mathcal{P}(W)$ such that:

    (a) $\langle s, E \rangle \in R_E(\langle s, E \rangle)$,

    (b) If $\langle s', E' \rangle \in R_E(\langle s, E \rangle)$, then $s' \in E$,

    (c) If $\langle s, E \rangle \in W$ and $s' \in E$, then $\langle s', E' \rangle \in R_E(\langle s, E \rangle)$ for some E'.

This definition generalizes normality structures in two ways. First, the mere fact that a state is compatible with a possible body of evidence doesn't mean

that the state is compatible with that being your total body of evidence. Second, evidential accessibility needn't be determined as a matter of having the same evidence. The first generalization is motivated by the idea that, when you are misperceiving, your evidence is different than it would be if things were as you perceived them to be. The second generalization is needed to accommodate the first: for example, to allow that when you are misperceiving it is compatible with your beliefs that you are perceiving. The Definition 2.3 of B remains unchanged, and the present generalization doesn't disrupt any of our propositions about the implications of (WEAK) STATISM, COMPARABILITY and COLLAPSE; the determination of generalized normality structures by an appropriately generalized probability structure goes through as before.

We will now describe a subclass of generalized normality structures that capture the particular kind of non-transparent evidence associated with misperception, misremembering, and similar cases familiar from recent discussions of skepticism. In a 'bad case' where you are misperceiving that you have hands (suppose you are a brain in a vat), we can ask what would count as a corresponding 'good case' is in which you are perceiving (and so have evidence that) you have hands. This will be a case where your evidence is stronger than in the bad case, *but no stronger than it needs to be* – i.e., it involves perception rather than illusion, but doesn't involve any further discoveries. Here is one strategy for making this idea precise:

Definition C.2. A *good/bad* structure is a generalized normality structure such that, if $\langle s', E' \rangle \in R_E(\langle s, E \rangle)$, then $E' = \bigcup \{E'' \subseteq E : \langle s', E'' \rangle \in W\}$.

Good/bad structures are still a generalization of normality structures. They have two notable properties that are not shared by all generalized normality structures. The first is that evidential accessibility is transitive – if $v \in R_E(w)$, then $R_E(v) \subseteq R_E(w)$. We think this is an idealization, but a useful one for isolating the evidential structure distinctive of skeptical scenarios. The second is that, for any situation and any state compatible with your evidence in that situation, there is a unique evidentially accessible situation in which that state obtains. This means that evidential accessibility is recoverable from W. This affords an alternative characterization of good/bad structures: start with a set W such that, whenever $\langle s, E \rangle \in W$ and $s' \in E$, $\langle s', \bigcup \{E'' \subseteq E : \langle s', E'' \rangle \in W\} \rangle \in W$; then define R_E from W in the obvious way.

This is only a brief and programmatic suggestion for how we might extend the present framework for theorizing about the dynamics of belief to cases where evidence is non-transparent. We mention it both as a direction for further research and to illustrate that the transparency of evidence is not an essential assumption of our framework for theorizing about belief dynamics. That is, non-transparent evidence is compatible with the two-component way of modelling situations in terms of which we have defined belief and learning about the state of the world (and also with the use of probability structures to determine normality relations, as explained in Goodman and Salow (2021, Appendix A)). That being said, we also think further developments of non-transparent models of belief dynamics are probably premature, because non-transparent evidence

raises challenging questions about how to think about *synchronic* belief in terms of evidential accessibility and comparative normality, for reasons explained in Goodman and Salow (in preparation). (In brief, the challenge arises in 'abnormal good cases', where the asymmetries of evidential accessibility and of comparative normality push in opposite directions.)

D   Other Formalisms

This appendix explains how normality structures can be used to interpret the $*$ operator used to formulate standard theories of belief revision; the relation $\mathrel{\vert\!\sim}$ of nonmonotonic consequence (conceived as a kind of evidential support); and the notion of conditional belief.

D.1   AGM

Theories of belief revision, like AGM, are usually formulated in terms of a binary operator $*$ that, given a set of sentences $\mathbf{A}$ (your original beliefs) and a sentence φ (what you learn), outputs another set of sentences $\mathbf{A}*\varphi$ (your updated beliefs).

Consider, for example, the following theorem of AGM:

PRESERVATION: If $\neg\varphi \notin \mathbf{A}$, then $\mathbf{A} \subseteq \mathbf{A} * \varphi$.

We can reinterpret this principle in our framework by replacing $\mathbf{A}$ and $\mathbf{A}*\varphi$ with $\{p \subseteq S : B(S) \subseteq p\}$ and $\{p \subseteq S : B(\varphi) \subseteq p\}$, respectively, replacing Boolean connectives in φ with corresponding set-theoretic operations, and restricting our attention to *full* normality structures in which $\mathcal{E} = \mathcal{P}(S)\backslash\{\emptyset\}$ (so that $*$ always is well-defined). PRESERVATION is then equivalent to $\Diamond-$.

Similarly, as advertised in §1, $\Diamond R$, $\Box+$, $\Box-$, and $\Box R$ are also equivalent to theorems of AGM recast in the present framework. We saw earlier that all of those principles are valid in the class of normality structures satisfying STATISM, COLLAPSE, and COMPARABILITY. In fact, this is true for all theorems of AGM:

Proposition 23. *Under the above translation of claims about $*$ into claims about the B operator, all theorems of AGM are valid on the class of full normality structures satisfying* STATISM, COLLAPSE, *and* COMPARABILITY.

This result is unsurprising given the formal parallel between such normality structures and models of AGM in terms of plausibility orders, originally introduced by Grove (1988). Simplifying slightly but inessentially, these models consist of a set of worlds equipped with a well-founded total preorder, understood as a relation of comparative plausibility. An agent believes what is true in the most plausible worlds, and after learning p believes what is true in the most plausible worlds in which p is true. Given STATISM we can see worlds as corresponding to our states; given COLLAPSE we can see comparative plausibility as our $\succeq$. An agent's initial beliefs being determined by the most plausible of *all* worlds corresponds to requiring that $S \in \mathcal{E}$; COMPARABILITY is then equivalent to the claim that the comparative plausibility preorder is total.

D.2   Nonmonotonic consequence

Closely related to belief revision is the large literature on nonmonotonic consequence relations, understood as relations of inductive support.[48] Within normality structures, we can interpret $p \mathrel{|\!\sim} q$ as the claim that, if p is your total evidence, then you believe q:[49]

Definition D.1. $\varphi \mathrel{|\!\sim} \psi := B(\varphi') \subseteq \psi'$, where φ' and ψ' result from replacing Boolean connectives in φ and ψ with corresponding set theoretic operations.

This definition allows us to use normality structures as a bridge between theories of belief revision and theories of nonmonotonic consequence. In particular, many of the principles discussed in the main text and in Appendix A then correspond to principles of nonmonotonic logic from Kraus et al. (1990), as per the following table:[50]

$\Diamond-$	RATIONAL MONOTONY	If $p \mathrel{|\!\sim} r$ and $p \mathrel{|\!\not\sim} \neg q$, then $p \wedge q \mathrel{|\!\sim} r$
$\Box+$	CUT	If $p \mathrel{|\!\sim} r$ and $p \mathrel{|\!\sim} q$, then $p \wedge q \mathrel{|\!\sim} r$
$\Box-$	CAUTIOUS MONOTONY	If $p \mathrel{|\!\sim} r$ and $p \mathrel{|\!\sim} q$, then $p \wedge q \mathrel{|\!\sim} r$
$\Pi+$	WEAK OR	If $p \wedge q \mathrel{|\!\sim} r$ and $p \wedge \neg q \mathrel{|\!\sim} r$, then $p \mathrel{|\!\sim} r$
$\Pi-$	NEGATION RATIONALITY	If $p \wedge q \mathrel{|\!\not\sim} r$ and $p \wedge \neg q \mathrel{|\!\not\sim} r$, then $p \mathrel{|\!\not\sim} r$
$C+$	OR	If $p \mathrel{|\!\sim} q$ and $q \mathrel{|\!\sim} r$ then $p \vee q \mathrel{|\!\sim} r$
$C-$	DISJUNCTION RATIONALITY	If $p \mathrel{|\!\not\sim} q$ and $q \mathrel{|\!\not\sim} r$ then $p \vee q \mathrel{|\!\not\sim} r$
FRONT-LOADING	S	If $p \wedge q \mathrel{|\!\sim} r$ then $p \mathrel{|\!\sim} q \supset r$

Shoham (1987) and especially Kraus et al. (1990) develop a 'preferential' semantics for $\mathrel{|\!\sim}$ that is related to our normality structures; roughly, the idea is that $p \mathrel{|\!\sim} q$ is true just in case $\{s \in p : \forall s' \in p(s' \not\gg s)\} \subseteq q$. In the simplest case, the only constraint on $\gg$ is that it be irreflexive, transitive, and well-founded; the resulting class of models validates the system **P** that includes CAUTIOUS MONOTONY and CUT but not RATIONAL MONOTONY. Lehmann and Magidor (1992) show that further requiring $\gg$ to be *modular* (if $x \gg y$, then for any z either $z \gg y$ or $x \gg z$) defines the class of such structures that validate RATIONAL MONOTONY as well.

These results parallel Propositions 1 and 2 as follows. The relata of $\gg$ correspond to our states rather than to our situations; so the fact that, in our system, $\Box-$ and $\Box+$ are validated given STATISM corresponds to the fact that CAUTIOUS MONOTONY and CUT are valid in the general class of preferential models. The basic preferential models build in no constraints corresponding to COMPARABILITY or COLLAPSE. So the invalidity of $\Diamond-$ in our models when these constraints are not imposed corresponds to the invalidity of RATIONAL MONOTONY in the class of all preferential models. And imposing COMPARABILITY and COLLAPSE

---

[48] Genin (2019) is an excellent entry point to this literature and its relation to belief revision.

[49] Compare Kraus et al. (1990), Stalnaker (1994) and Smith (2018b). Relatedly, Makinson and Gärdenfors (1991) interpret $p \mathrel{|\!\sim} q$ as $q \in \mathbf{A} * p$, for some salient background theory **A**.

[50] Strictly speaking, in the case of $\Pi+$, $\Pi-$, $C+$, and $C-$, the relevant principle of nonmonotonic logic corresponds to the special case where Π or $\mathcal{C}$ contain exactly two events.

entails that $\gg$ is modular on evidentially accessible situations.[51] So the fact that $\Diamond-$ becomes valid when we impose these constraints on normality structures corresponds to the fact that RATIONAL MONOTONY is valid on class of modular preferential models.

For a more systematic correspondence, let system $\mathbf{N}$ be axiomatized by the following principles:

IDENTITY	$p \mathrel{\vdash\hspace{-0.6em}\sim} p$
LEFT LOGICAL EQUIVALENCE	If $p \mathrel{\vdash\hspace{-0.6em}\sim} r$ and $\models p \leftrightarrow q$, then $q \mathrel{\vdash\hspace{-0.6em}\sim} r$
RIGHT WEAKENING	If $p \mathrel{\vdash\hspace{-0.6em}\sim} q$ and $q \models r$, then $p \mathrel{\vdash\hspace{-0.6em}\sim} r$
AND	If $p \mathrel{\vdash\hspace{-0.6em}\sim} q$ and $p \mathrel{\vdash\hspace{-0.6em}\sim} r$, then $p \mathrel{\vdash\hspace{-0.6em}\sim} q \wedge r$

The standard System $\mathbf{P}$ is then the result of adding OR, CAUTIOUS MONOTONY and CUT to $\mathbf{N}$; and the standard system $\mathbf{R}$ is the result of further adding RATIONAL MONOTONY to $\mathbf{P}$.[52] We then have the following result:

Proposition 24.

- *$\mathbf{R}$ is valid on the class of normality structures satisfying* STATISM, COMPARABILITY *and* COLLAPSE.

- *$\mathbf{P}$ is valid on the class of normality structures satisfying* STATISM.

- *$\mathbf{N}$ is valid on the class of all normality structures.*

Remark D.1. This result depends on the fact that $\mathrel{\vdash\hspace{-0.6em}\sim}$ is defined in terms of B, which in turn *presupposes* that the relevant event is a possible body of evidence. A parallel result does not hold for the relation $p \mathrel{\vdash\hspace{-0.6em}\sim}' q$, defined to mean that either $p \notin \mathcal{E}$ or $B(p) \subseteq q$. For example, consider the following theorem of $\mathbf{P}$:

($\dagger$) If $p \mathrel{\vdash\hspace{-0.6em}\sim} \bot$, then $p \wedge q \mathrel{\vdash\hspace{-0.6em}\sim} \bot$.

This claim is valid on the class of normality structures: it presupposes that $\mathcal{E}$ contains both p and $p \cap q$, and it is true in all structures satisfying this presupposition. By contrast, the parallel principle with $\mathrel{\vdash\hspace{-0.6em}\sim}'$ in place of $\mathrel{\vdash\hspace{-0.6em}\sim}$ fails in normality structures where $p \notin \mathcal{E}$ but $\emptyset \neq p \cap q \in \mathcal{E}$.

Relatedly, although Proposition 24 can be strengthened to a completeness result for $\mathbf{R}$ and $\mathbf{P}$ (see Lehmann and Magidor (1992) and Kraus et al. (1990), respectively), the same cannot be said for $\mathbf{N}$, since ($\dagger$) is not a theorem of $\mathbf{N}$.

D.3 Conditional Belief

Closely connected to the topic of belief revision is the topic of conditional belief. In fact, it is tempting to understand the latter in terms of the former: you believe

---

[51]Assume COMPARABILITY and COLLAPSE. Suppose $x \gg y$. If $z \succeq x$, then $z \gg y$, by the condition on $\succeq$ and $\gg$ from Definition 2.1.5b and the reflexivity of $\succeq$. And if $z \not\succeq x$, then $x \succeq z$, by COMPARABILITY, and hence $x \gg z$, by COLLAPSE. So $\gg$ is modular.

[52]These axiomatizations are redundant: CUT and AND are interderivable given the other axioms of $\mathbf{P}$ and CAUTIOUS MONOTONY is redundant in $\mathbf{R}$ given RATIONAL MONOTONY.

Smith (2016, 2018b) defends $\mathbf{R}$ as a theory of propositional justification (understood as a kind of evidential support which is in turn understood as a kind of nonmonotonic consequence).

p conditional on q if and only if you would believe p after learning q. However, such an account faces the problem that we can make sense of belief conditional on q even when q is an event that you couldn't possibly learn: for example, events corresponding to claims of the form 'p but I'll never learn that p'; see also Appendix A.

This motivates a different approach. Belief conditional on q can instead be characterized exactly like belief – in terms of comparative normality and evidential accessibility – except with evidential accessibility now restricted to situations in which q obtains. More precisely:

Definition D.2. Given evidence E, you believe p conditional on q just in case $B^q(E) \subseteq p$, where:

$R_E^q(w) = \{\langle s, E \rangle \in R_E(w) : s \in q\}$;

$R_B^q(w) = \{v \in R_E^q(w) : \forall u \in R_E^q(w)(u \gg v)\}$;

$B^q(E) = \{s' : \langle s', E' \rangle \in R_B^q(\langle s, E \rangle) \text{ for some } s \in E \text{ and } E' \in \mathcal{E}\}$.

If STATISM holds and q is something you could learn, then you now believe p conditional on q if and only if you would believe p after discovering q. But this equivalence depends on STATISM, which we reject. In particular, in the models motivated in §7, having previously believed p conditional on q is neither necessary nor sufficient for believing p upon learning q. The failures of FRONT-LOADING discussed in §7 are counterexamples to the necessity direction. For a counterexample to the sufficiency direction, modify the **Flipping for Heads** thought experiment as discussed in Goodman and Salow (2023a, p.138): you believe that the coin was flipped only once (p) conditional on it being flipped either once or at least $n + 1$ times (q), but after learning q you don't believe p.

E   Proofs

Proof of **Proposition 1**:

WEAK STATISM $\Rightarrow \Diamond R$

Suppose that $B(E) \cap$p$\neq \emptyset$. We assume WEAK STATISM and show that $B(E) \cap B(E \cap p) \neq \emptyset$.

Take any $s_1 \in B(E) \cap p$. If $s_1 \in B(E \cap p)$, we are done. So suppose $s_1 \notin B(E \cap p)$. Then there must be some $s \in E \cap p$ such that $s \gg_{E \cap p} s_1$. Moreover, at least one such s – call it s_2 – must be in $B(E \cap p)$: otherwise there would, for each such s, be another such s' with $s' \gg_{E \cap p} s$, violating the well-foundedness of $\gg$. Since $s_2 \gg_{E \cap p} s_1$, $s_2 \succeq_{E \cap p} s_1$, and thus by WEAK STATISM, $s_2 \succeq_E s_1$. So, since $s_1 \in B(E)$, $s_2 \in B(E)$ as well. So $s_2 \in B(E)$ and $s_2 \in B(E \cap p)$, showing that $B(E) \cap B(E \cap p) \neq \emptyset$.

STATISM $\Rightarrow \Box+$ and $\Box-$

Suppose that $B(E) \subseteq p$. We assume STATISM and show that $B(E \cap p) = B(E)$.

Suppose first that $s_1 \in B(E \cap p)$, i.e. $s_1 \in E \cap p$ and there is no $s_2 \in E \cap p$ such that $s_2 \gg_{E \cap p} s_1$. Since $B(E) \subseteq E \cap p$, this means that there is no $s_2 \in B(E)$ such

that $s_2 \gg_{E\cap p} s_1$. By STATISM, there is then no $s_2 \in B(E)$ such that $s_2 \gg_E s_1$. So $s_1 \in B(E)$. So $B(E \cap p) \subseteq B(E)$.

Suppose instead that $s_1 \notin B(E \cap p)$. Then either $s_1 \notin E \cap p$ or there is an $s_2 \in E \cap p$ such that $s_2 \gg_{E\cap p} s_1$. If the former $s_1 \notin B(E)$, since $B(E) \subseteq E \cap p$. If the latter, then, by STATISM, $s_2 \gg_E s_1$, and so again $s_1 \notin B(E)$. So $B(E) \subseteq B(E \cap p)$.

<u>STATISM, COLLAPSE, and COMPARABILITY $\Rightarrow \Diamond-$</u>

Suppose that $B(E) \cap \mathrm{p} \neq \emptyset$. We assume STATISM, COLLAPSE, and COMPARABILITY, and show that $B(E \cap p) \subseteq B(E)$.

Suppose that $s_1 \in E$ but $s_1 \notin B(E)$, i.e. there is some $s_2 \in E$ with $s_2 \gg_E s_1$. Let $s_3 \in B(E) \cap p$. By COMPARABILITY, either $s_3 \succeq_E s_2$ or $s_2 \succ_E s_3$. If $s_2 \succ_E s_3$ then, by COLLAPSE, $s_2 \gg_E s_3$, contradicting $s_3 \in B(E)$. So $s_3 \succeq_E s_2$. So $s_3 \gg_E s_1$. So, by STATISM, $s_3 \gg_{E\cap p} s_1$. So $s_1 \notin B(E \cap p)$. So $B(E \cap p) \subseteq B(E)$. $\square$

Proposition 2: $\Diamond-$ fails in the model of **Three Friends** in §4 (which validates STATISM and COLLAPSE) and in the model of **Flipping for Heads** in §5 (which validates STATISM and COMPARABILITY). $\Box+$ fails in the modified model of **Flipping for Heads** in §7 (in which WEAK STATISM and COMPARABILITY hold). For a model in which $\Box-$ fails and WEAK STATISM holds: $S = \{a, b, c\}$; $\mathcal{E} = \{S, \{a, b\}\}$; $a \succ_S b, c$ and $a \succ_{\{a,b\}} b$ (satisfying WEAK STATISM); $a \gg_S b, c$ and $a \not\gg_{\{a,b\}} b$. $B(S) \subseteq \{a, b\}$, but $\{a, b\} = B(S \cap \{a, b\}) \not\subseteq B(S) = \{a\}$. For a model in which $\Box R$ fails and COMPARABILITY and COLLAPSE hold, modify the previous model by having $b \gg_{\{a,b\}} a$ with the obvious adjustments: $B(S)$ is unchanged, but $B(\{a, b\}) = \{b\}$, so $B(S) \cap B(S \cap \{a, b\}) = \emptyset$.

Proof of **Proposition 3**: <u>COMPARABILITY and STATISM $\Rightarrow \Pi-$</u>

Let $\Pi = \{p_1, \ldots, p_n\}$ be a finite partition of E. We assume that $B(E \cap p_i) \not\subseteq B(E)$ for every i, and show that this assumption leads to contradiction.

By assumption, there is, for each i, some $s_1^i \in B(E \cap p_i) \backslash B(E)$. Since $s_1^i \notin B(E)$, there is a corresponding $s_2^i \in E$ such that $s_2^i \gg_E s_1^i$. Moreover, $s_2^i \geq_E s_3^i$ for any $s_3^i \in E \cap p_i$. For, by COMPARABILITY, either $s_2^i \geq_E s_3^i$ or $s_3^i \geq_E s_2^i$; but if $s_3^i \geq_E s_2^i$ then $s_3^i \gg_E s_1^i$, so by STATISM $s_3^i \gg_{E\cap p_i} s_1^i$, contradicting $s_1^i \in B(E \cap p_i)$.

Now consider $\{s_2^i : i \leq n\}$. This is a finite set, so it contains a 'minimal' s, i.e. an s such that $s' \not\succ_E s$ for any other member s'. By COMPARABILITY, this means that $s \geq_E s_2^i$ for every i. Now, $s \in p_j$ for some j. But then $B(E \cap p_j) \subseteq B(E)$ after all. For suppose that $s' \in E$ but $s' \notin B(E)$, so that there is some $s'' \in E$ with $s'' \gg_E s'$. $s'' \in p_i$ for some i, so, by the above, $s_2^i \geq_E s''$. But then $s \geq_E s_2^i \geq_E s'' \gg_E s'$. So $s \gg_E s'$. So by STATISM, either $s' \notin p_j$ or $s \gg_{E\cap p_j} s'$. Either way, $s' \notin B(E \cap p_j)$. So $B(E \cap p_j) \subseteq B(E)$, contradicting our assumption. $\square$

Remark on Proposition 3: To see why Π must be finite, consider the following

model (specifying $\geq$ and $\gg$ directly on states, thus satisfying STATISM):

- $S = \{\langle x, y \rangle : x \in \mathbb{N}, y \in \{0, 1\}\}$.

- $\langle x, y \rangle \geq \langle u, v \rangle$ just in case either $y > v$ or $y = v$ and $x \geq u$.

- $\langle x, y \rangle \gg \langle u, v \rangle$ just in case both $y > v$ and $x > u$.

- $\mathcal{E} = \{S\} \cup \{E_n = \{\langle x, 0 \rangle, \langle x, 1 \rangle\} : n \in \mathbb{N}\}$.

For each n, $B(E_n) = \{\langle n, 0 \rangle, \langle n, 1 \rangle\} \nsubseteq \{\langle x, 1 \rangle : x \in \mathbb{N}\} = B(S)$. Yet the model satisfies STATISM and COMPARABILITY, and $\{E_n\}$ is a (infinite) partition of S. (By contrast, the proof of Proposition 7 doesn't require Π to be finite, since $\succ$ is well-founded in normality structures derived from simple probability structures.)

Proof of **Proposition 4**: Failure in a normality structure satisfying STATISM implies failure in a normality structure satisfying STATISM and COLLAPSE.
Consider any normality structure $\langle S, \mathcal{E}, W, \succeq, \gg \rangle$ satisfying STATISM. Let $\gg^*$ be the reflexive closure of $\gg$. Then $\langle S, \mathcal{E}, W, \gg^*, \gg \rangle$ is a normality structure satisfying COLLAPSE that agrees with the original structure about B. It thus invalidates any principle invalidated by the original. □

The proofs of **Propositions 5 and 6** are routine.

Proof of **Proposition 7**:

WEAK STATISM holds in normality structures derived from simple probability structures
If $s, s' \in E$ and $s, s' \in E'$, and $Pr(E), Pr(E') > 0$ then either $\frac{Pr(\{s\}|E)}{Pr(\{s'\}|E)} = \frac{Pr(\{s\}|E')}{Pr(\{s'\}|E')}$ or $Pr(\{s'\}) = 0$. Either way, $Pr(\{s\}|E) \geq Pr(\{s'\}|E)$ iff $Pr(\{s\}|E') \geq Pr(\{s'\}|E')$.

$\square-$ holds in normality structures derived from simple probability structures.
Immediate from Proposition 11 below.

$\Pi-$ holds in normality structures derived from simple probability structures.
Suppose $\langle S, \mathcal{E}, W, \succeq, \gg \rangle$ is determined by $\langle S, \mathcal{E}, W, Pr, t \rangle$. Since WEAK STATISM holds, we may treat $\succeq$ as a relation on states. We write $\succ s$ for $\{s' : s' \succ s\}$, and observe that $s \in B(E)$ if and only if $Pr_E(\succ s) < t$. Note that $\succ$ is well-founded in normality structures determined from probability structures.

Suppose for contradiction that $\Pi-$ fails: $B(E \cap p_i) \backslash B(E) \neq \emptyset$ for each $p_i \in \Pi$. For each i, let s_i be some member of $B(E \cap p_i) \backslash B(E)$, and k be such that $s_k \succeq s_i$ for all i (which is possible by COMPARABILITY and the well-foundedness of $\succ$). Since $s_i \in B(E \cap p_i)$, $Pr_{E \cap p_i}(\succ s_i) < t$. Since $\succ s_k \subseteq \succ s_i$ for all i, we have $Pr_{E \cap p_i}(\succ s_k) \leq Pr_{E \cap p_i}(\succ s_i) < t$. So, by total probability, $Pr_E(\succ s_k) < t$. So $s_k \in B(E)$, contradicting our assumption that $s_k \in B(E \cap p_k) \backslash B(E)$.
□

Proposition 8: as explained in the main text, STATISM and $\square+$ fail in models of **Flipping for Heads** in which you are told that your initial belief was true.

The proofs of **Propositions 9** and **10** are routine.

Proof of **Proposition 11**: $\Box-$ holds in all normality structures derived from probability structures.

Suppose $\langle S, \mathcal{E}, W, \succeq, \gg \rangle$ is determined by $\langle S, \mathcal{E}, W, Q, Pr, t \rangle$. To establish $\Box-$, we suppose that $B(E) \subseteq p$, and show that it follows that $B(E \cap p) \subseteq B(E)$.

Note that if $s_2 \in [s_1]_Q \cap E$, then $s_2 \succeq_E s_1$. So if $s_1 \in B(E)$, $[s_1]_Q \cap E \subseteq B(E) \subseteq p$. So for any s_3, if $Pr_E([s_1]_Q) \geq Pr_E([s_3]_Q)$, then also $Pr_{E \cap p}([s_1]_Q) \geq Pr_{E \cap p}([s_3]_Q)$. So if $s_3 \succeq_{E \cap p} s_1$ and $s_1 \in B(E)$, then $s_3 \succeq_E s_1$ and $s_3 \in B(E)$.

Moreover, since $B(E) \subseteq p$, $Pr_{E \cap p}(B(E)) \geq Pr_E(B(E)) \geq t$.

Observe that $B(E \cap p)$ is the minimal $X \subseteq E \cap p$ such that (i) if $s_1 \in X$ and $s_3 \succeq_{E \cap p} s_1$, then $s_3 \in X$, and (ii) $Pr_{E \cap p}(X) \geq t$. 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. $\qquad\square$

Proposition 12: as explained in the main text, WEAK STATISM and $\Diamond R$ can fail in models of **Bias Detection**, and $\Pi-$ in models of **Celebrity Hike**.

Proposition 13: routine (see also Hawthorne (1996) and Shear and Fitelson (2019)). Note that we interpret the relevant principles using their English formulations, not in terms of the B operator (which doesn't make sense in the Lockean setting, since belief isn't closed under conjunction).

Proposition 14 is proved in footnote 6.

Proposition 15 follows from Theorem 7 in Leitgeb (2017, p.121).

Proof of **Proposition 16**:

Topical structures that satisfy t-STABILITY satisfy $\Diamond-$

Suppose that $B(E) \cap p \neq \emptyset$ and $E \cap p \in \mathcal{E}$. By topicality, $E \cap p$ is topical. So, by t-STABILITY, $Pr_{E \cap p}(B(E)) \geq t$. Moreover, topical structures satisfy ORTHOGONALITY, so WEAK STATISM holds, so $B(E) \cap p$ is closed under $\succeq_{E \cap p}$ (since $B(E)$ is always closed under $\succeq_E$). Since $B(E \cap p)$ is the minimal subset of $E \cap p$ that both has above threshold probability given $E \cap p$ and is closed under $\succeq_{E \cap p}$, it follows that $B(E \cap p) \subseteq B(E)$.

Topical structures that satisfy $\Diamond-$ satisfy t-STABILITY

Suppose that $B(E)$ is not weakly t-stable, so that for some topical p consistent with $B(E)$, $t > Pr_E(B(E)|p) = Pr_{E \cap p}(B(E))$. By topicality, $E \cap p \in \mathcal{E}$. Since $B(E \cap p)$ must be such that $Pr_{E \cap p}(B(E \cap p)) \geq t$, we have $B(E \cap p) \not\subseteq B(E)$, violating $\Diamond-$.

Proposition 17 is immediate from definitions.

Proposition 18: The claims about validities follow from Propositions 1 and 17. The claims about invalidities are established by the counterexamples in the main text (since ΠR-failures are both $\Pi-$ and $\Pi+$ failures).

Proof sketch of **Proposition 19**: Normality structures Levi-determined by probability structures satisfying ORTHOGONALITY satisfy COMPARABILITY and WEAK STATISM, and validate $\Box-$, $\Pi-$, and $\Diamond R$.

WEAK STATISM and COMPARABILITY are routine, and $\Diamond R$ follows by Proposition 1. $\Box-$ holds, roughly, because removing some sufficiently-below-average-probability states from your evidence cannot make the remaining previously sufficiently-below-average-probability states no longer such. $\Pi-$ holds because a collection of disjoint sets cannot each have a higher average probability than its union does.

Proof of **Proposition 20**: $\Box+$ can fail in normality structures Levi-determined by probability structures satisfying ORTHOGONALITY.

Let $S = \{a, b, c\}$, $\mathcal{E} = \{S, \{a, b\}\}$, $Q = \{\{a\}, \{b\}, \{c\}\}$, and $Pr(\{a\}) = 1 - x - \epsilon > Pr(\{b\}) = x > Pr(\{c\}) = \epsilon \approx 0$, where $\frac{t}{3} < x < \frac{t}{2}$. In the Levi-determined normality structure, $B(E) = \{a, b\}$ but $B(\{a, b\}) = \{a\}$. $\qquad\square$

Proof of **Proposition 21:**

$\Box-$ can fail in normality structures Levi-determined by probability structures.

Let $S = \{s_1, \ldots s_{1100}\}$; Pr uniform; $Q = \{p, p', r_1, \ldots r_{98}\}$, where p contains 110 states while p' and each r_i contain 10 states. Suppose $t > \frac{10}{11}$. Then $B(S) = p$. Now suppose E contains all members of p and of p', and one member from each r_i. Since E doesn't rule out any answers to Q, $|\{q \in Q : q \cap E \neq \emptyset\}| = |Q| = 100$. And $Pr_E(p') = \frac{10}{218} > \frac{t}{100}$. So $p' \subseteq B(E)$. So learning E, which you already believed, results in you giving up the belief that p' does not obtain.

$\Diamond R$ and ΠR can fail in normality structures Levi-determined by probability structures.

ΠR fails in **Drawing a Card** for the same reason as it does for LK-determination. Since one of the members of the partition was compatible with your initial beliefs, $\Diamond R$ fails as well. $\qquad\square$

Proposition 22 is proved in the main text.

Proposition 23 is an immediate corollary of Grove (1988), given the parallels described.

Proposition 24: for **P** and **R**, immediate from the parallel between the relevant class of normality structures and the models of **P** and **R** in Kraus et al. (1990) and Lehmann and Magidor (1992). The proof for **N** is routine.

References

Carlos Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50:510–530, 1985.

Andrew Bacon. Giving your knowledge half a chance. *Philosophical Studies*, 171:373–397, 2014.

Alexandru Baltag and Sonja Smets. A qualitative theory of dynamic interactive belief revision. In Giacomo Bonanno, Wiebe van der Hoek, and Michael Wooldridge, editors, *Logic and the Foundations of Game and Decision Theory*. Amsterdam University Press, 2008.

Bob Beddor and Carlotta Pavese. Modal virtue epistemology. *Philosophy and Phenomenological Research*, 101:61–79, 2020.

Craig Boutilier. Iterated revision and minimal change of conditional beliefs. *Journal of Philosophical Logic*, 25:263–305, 1996.

John Cantwell and Hans Rott. Probability, coherent belief and coherent belief changes. *Annals of Mathematics and Artificial Intelligence*, 87(3):259–291, 2019.

Sam Carter and John Hawthorne. Dogmatism & inquiry. *Mind*, forthcoming.

David Chalmers. *Constructing the World*. Oxford University Press, 2012.

Andrew Dabrowski, Lawrence Moss, and Rohit Parikh. Topological reasoning and the logic of knowledge. *Annals of Pure and Applied Logic*, 78:73–110, 1996.

Adnan Darwiche and Judea Pearl. On the logic of iterated belief revision. *Artificial intelligence*, 89:1–29, 1997.

Nilanjan Das. The value of biased information. *British Journal for the Philosophy of Science*, 74(1):25–55, 2023.

Cian Dorr, Jeremy Goodman, and John Hawthorne. Knowing against the odds. *Philosophical Studies*, 170:277–287, 2014.

Cian Dorr, Jacob Nebel, and Jake Zuehl. The case for comparability. *Noûs*, 57 (2):414–453, 2023.

Kevin Dorst. Evidence: A guide for the uncertain. *Philosophy and Phenomenological Research*, 100:586–632, 2020.

Kevin Dorst and Matthew Mandelkern. Good guesses. *Philosophy and Phenomenological Research*, 3(105):581–618, 2023.

Igor Douven and Hans Rott. From probabilities to categorical beliefs: Going beyond toy models. *Journal of Logic and Computation*, 28(6):1099–1124, 2018.

Richard Foley. *Working without a net: A study of egocentric epistemology.* Oxford University Press, 1993.

Konstantin Genin. Full and partial belief. In Richard Pettigrew and Jonathan Weisberg, editors, *The Open Handbook of Formal Epistemology.* The PhilPapers Foundation, 2019.

Simon Goldstein and John Hawthorne. Knowledge from multiple experiences. *Philosophical Studies*, 179:1341–72, 2022.

Jeremy Goodman. Degrees of confidence are not subjective probabilities. In *Proceedings of Sinn und Bedeutung 28*, 2024.

Jeremy Goodman and Ben Holguín. Thinking and being sure. *Philosophy and Phenomenological Research*, 3(106):634–54,, 2023.

Jeremy Goodman and Bernhard Salow. Taking a chance on KK. *Philosophical Studies*, 175:183–196, 2018.

Jeremy Goodman and Bernhard Salow. Knowledge from probability. In Joseph Y. Halpern and Andres Perea, editors, *Theoretical Aspects of Rationality and Knowledge 2021 (TARK 2021)*, pages 171–186, 2021.

Jeremy Goodman and Bernhard Salow. Epistemology normalized. *Philosophical Review*, 132, 2023a.

Jeremy Goodman and Bernhard Salow. Belief revision from probability. In Rineke Verbrugge, editor, *Theoretical Aspects of Rationality and Knowledge 2023 (TARK 2023)*, pages 308–317, 2023b.

Jeremy Goodman and Bernhard Salow. Paper on belief when partitionality fails. in preparation.

Adam Grove. Two modellings for theory change. *Journal of Philosophical Logic*, 17:157–170, 1988.

Ian Hacking. Isaac Levi, "Gambling with Truth: An Essay on Induction and the Aims of Science" (book review). *Synthese*, 17:444–448, 1967.

Ned Hall. How to set a surprise exam. *Mind*, 108:647–703, 1999.

Gilbert Harman. *Change in View: Principles of Reasoning.* MIT Press, 1986.

James Hawthorne. On the logic of nonmonotonic conditionals and conditional probabilities. *Journal of Philosophical Logic*, 25:185–218, 1996.

John Hawthorne. Deeply contingent a priori knowledge. *Philosophy and Phenomenological Research*, 65(2):247–269, 2002.

Jaakko Hintikka. *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Ithaca: Cornell University Press, 1962.

Daniel Hoek. Minimal rationality and the web of questions. In Dirk Kindermann, Peter van Elswyk, and Andy Egan, editors, *Unstructured Content*. Oxford University Press, forthcoming.

Ben Holguín. Thinking, guessing, and believing. *Philosophers' Imprint*, 22(1): 1–34, 2022.

Frank Hong. *Uncertain Knowledge*. PhD thesis, University of Southern California, 2023.

Richard C Jeffrey. *The logic of decision*. University of Chicago Press, 1965.

Kevin Kelly and Hanti Lin. Beliefs, probabilities, and their coherent correspondence. In *Lotteries, Knowledge, and Rational Belief: Essays on the Lottery Paradox*, pages 185–222. Cambridge University Press, 2021.

Matthew Kotzen. The Bayesian and classical approaches to statistical inference. *Philosophy Compass*, 17(9), 2022.

Sarit Kraus, Daniel Lehmann, and Menachem Magidor. Nonmonotonic reasoning, preferential models, and cumulative logics. *Artificial Intelligence*, 44: 167–207, 1990.

John Kruschke. *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press/Elsevier, 2014.

Daniel Lehmann and Menachem Magidor. What does a conditional knowledge base entail? *Artificial intelligence*, 55:1–60, 1992.

Hannes Leitgeb. *The Stability of Belief: How Rational Belief Coheres with Probability*. Oxford University Press, 2017.

Hannes Leitgeb and Krister Segerberg. Dynamic doxastic logic: why, how, and where to? *Synthese*, 155:167–190, 2007.

Isaac Levi. *Gambling with Truth: An Essay on Induction and the Aims of Science*. MIT Press, 1967.

Isaac Levi. *For the sake of the argument: Ramsey test conditionals, inductive inference and nonmonotonic reasoning*. Cambridge University Press, 1996.

David Lewis. Attitudes *de dicto* and *de se*. *Philosophical Review*, 88:513–543, 1979.

David Lewis. Elusive knowledge. *Australasian Journal of Philosophy*, 74:549–567, 1996.

Hanti Lin. Belief revision theory. In Richard Pettigrew and Jonathan Weisberg, editors, *The Open Handbook of Formal Epistemology*. The PhilPapers Foundation, 2019.

Hanti Lin and Kevin Kelly. Propositional reasoning that tracks probabilisic reasoning. *Journal of Philosophical Logic*, 41:957–981, 2012.

David Makinson and Peter Gärdenfors. Relations between the logic of theory change and nonmonotonic logic. In André Fuhrmann and Michael Mourreau, editors, *The logic of theory change*. Springer, 1991.

Joshua Pearson. Belief revision revised. ms.

Hans Rott. A counterexample to six fundamental principles of belief formation. *Synthese*, 139(2):225–240, 2004.

Hans Rott. Stability and scepticism in the modelling of doxastic states: Probabilities and plain beliefs. *Minds and Machines*, 27:167–197, 2017.

Ted Shear and Branden Fitelson. Two approaches to belief-revision. *Erkenntnis*, 84:487–518, 2019.

Yoav Shoham. A semantical approach to nonmonotonic logics. *Proceedings of 2nd Symposium on Logic in Computer Science*, pages 275–279, 1987.

Martin Smith. *Between Probability and Certainty: What Justifies Belief*. Oxford UP, 2016.

Martin Smith. Coin trials. *Canadian Journal of Philosophy*, 48:726–741, 2018a.

Martin Smith. The logic of epistemic justification. *Synthese*, 195:3857–3875, 2018b.

Wolfgang Spohn. Ordinal conditional functions: A dynamic theory of epistemic states. In *Causation in decision, belief change, and statistics*, pages 105–134. Springer, 1988.

Robert Stalnaker. What is a non-monotonic consequence relation? *Fundamenta Informaticae*, 21:7–21, 1994.

Robert Stalnaker. On logics of knowledge and belief. *Philosophical Studies*, 128:169–99, 2006.

Robert Stalnaker. Iterated belief revision. *Erkenntnis*, 70(2):189–209, 2009.

Johan van Benthem. Dynamic logic for belief revision. *Journal of Applied Non-Classical Logics*, 17:129–155, 2007.

Minkyung Wang. *Aggregating Credences into a Belief*. PhD thesis, Ludwig-Maximilians-Universität München, 2022.

Roger White. Problems for dogmatism. *Philosophical Studies*, 131:525–557, 2006.

Timothy Williamson. *Knowledge and its Limits*. Oxford UP, 2000.

Timothy Williamson. Evidence of evidence in epistemic logic. In *Higher-Order Evidence: New Essays*. Oxford UP, 2019.

Seth Yalcin. Belief as question-sensitive. *Philosophy and Phenomenological Research*, 97:23–47, 2018.