Awareness growth and belief revision

Joe Roussos (joe.roussos@iffs.se)

June 15, 2020

Abstract

The problem of awareness growth, also known as the problem of new hypotheses, is a persistent challenge to Bayesian theories of rational belief and decision making. Cases of awareness growth include coming to consider a completely new possibilty (called expansion), or coming to consider finer distinctions through the introduction of a new partition (called refinement). Recent work has centred on Reverse Bayesianism, a proposal for rational awareness growth due to Karni and Vierø. This essay develops a "Reserve Bayesian" position and defends it against two challenges. The first, due to Anna Mahtani, says that Reverse Bayesian approaches yield the wrong result in cases where the growth of awareness constitutes an expansion relative to one partition, but a refinement relative to a different partition. The second, due to Steele and Stefánsson, says that Reverse Bayesian approaches cannot deal with new propositions that are evidentially relevant to old propositions. I argue that these challenges confuse questions of belief revision with questions of awareness change. Mahtani's cases reveal that the change of awareness itself requires a model, which specifies how propositions in the agent's old algebra are identified with propositions in the new algebra. I introduce a lattice-theoretic model for this purpose, which resolves Mahtani's problem cases and some of Steele and Stefánsson's cases. Applying my model of awareness change, then Reverse Bayesianism, and then a generalised belief revision procedure, resolves Steele and Stefánsson's remaining cases. In demonstrating this, I introduce a simple and general model of belief revision in the face of new information about previously unknown propositions.

1 Introduction

Bayesianism is a broad collection of approaches to decision theory and epistemology, centred on a commitment to representing beliefs probabilistically. It is a rich and successful theory (if so broad a church can be called a "theory"), both in philosophy and statistics. It faces some famous challenges, however. It assumes agents are logically omniscient, that they perform (or are representable as performing) complex probabilistic calculations beyond ordinary human capability, and that they have prior attitudes towards every possibility they may encounter. This essay is about the last of these problems for Bayesianism. Real agents are unaware of many possibilities, and when they encounter them for the first time it seems unreasonable to insist that, contrary to appearances, they do in fact have prior beliefs about them.

I will set out a model of growing awareness that specifies how an agent's probabilistic beliefs ought to be extended to cover a new space of possibilities and which provides constraints for the formation of new beliefs about the new possibilities. This problem has been discussed previously as the "problem of new hypotheses" in the philosophy of science (it is linked to the "problem of old evidence" raised by Glymour (1980), and has been discussed by Shimony (1970), Eels (1985), Earman (1992), and more recently Wenmackers and Romeijn (2016)). My attention will be on a more recent literature in decision theory and epistemology, where the problem has come to be known as "awareness growth".

Our starting point is a probabilistic model of rational belief. A probabilistic model of belief requires at least two components: an algebra of propositions and a probability function. It is usual to talk about the probability function as representing the agent's beliefs, or degrees of belief. Awareness, as I will think of it, is a necessary condition for taking an attitude toward a proposition. As the probability function represents the attitude (in this case belief), it is a natural suggestion to use the algebra itself to model the agent's awareness. At first pass we will say that a proposition within the algebra is one that the agent is aware of, and any propositions not in the algebra are those the agent is unaware of. The algebra also encodes logical relationships between propositions. Changes to awareness will be thought of as, in the first instance, changes to the algebra. (This sets up a helpful parallel between the problem of probabilistic belief extension and revision in the face of new possibilities, and familiar techniques of qualitative belief revision pioneered by Levi (1977, 1980), Alchourrón, Gärdenfors, and Makinson (1985).) I will idealise and assume that the agent is logically omniscient: their degrees of belief reflect all the deductive logical relationships between propositions of which they are aware.[1]

As the probability function is defined on the algebra, this modelling choice means that in our model agents have no attitudes to propositions they are unaware of. This, I claim, is what we want. I'm interested in what we might call *true unawareness*, and true awareness growth: situations in which agents confront genuinely new possibilities. The common cases of this include learning a new concept, considering a new scientific theory, learning things about a city you've never visited before, and many others.

This is a narrower definition of awareness growth than others in the literature have used. Bradley, Steele and Stefánsson consider cases in which agents have forgotten or failed to consider a proposition, or in which they deliberately exclude a proposition from consideration, as cases of unawareness and treat remembering or considering as awareness growth. My restriction to true awareness growth is less severe than it may seem. While agents who have, for example, forgotten something are not truly unaware of the relevant propositions, they may be fruitfully *modelled as if they were* for the purposes of an analysis. We might use a model of unawareness to study the rational constraints on forgetting, and a model of awareness growth to study remembering. Nevertheless, we can recognise a significant conceptual difference between an agent who is truly unaware

---

[1]This is an important idealisation. I am not attempting to solve all the problems of Bayesianism at once; rather I am isolating awareness growth as a particular challenge, and showing how it can be resolved. With this idealisation in place, we can make use of a comment by Isaac Levi (1991) to illuminate the nature of awareness. Just as Levi points out that a logically closed belief set contains the sentences an agent is committed to believe, the awareness state contains propositions that the agent is committed to take some credal attitude to.

of a proposition—such as a scientist in the moments just prior to first learning of a new theory—and an agent who is temporarily excluding something from consideration.

A complication in this discussion is that the literature operates in terms of toy models involving small sets of prosaic propositions, even when discussing true unawareness. It is hard to shake the intuition that surely the agents know something about the "new" propositions in these stories—*obviously they know about buses going to town*, or *surely they had heard of French films before*. This is a problem with the examples, and with the intuitions attached to them, when it comes to studying true unawareness. This is further compounded by modelling cases where the agent does know about the bus going to town, *but has forgotten it*, as unawareness. I shall pick my way through this thicket by considering only true unawareness, and forcefully rejecting those hangover intuitions in toy examples.

The general problem that I consider is how an agent's beliefs should rationally change when they become aware new propositions. The proposal I defend is as follows. An agent's initial awareness state is represented by the Boolean algebra on which their probabilities are defined. After their awareness grows, their beliefs are again represented by probabilities defined on an algebra. The old algebra is related to the new by a kind of mapping called a lattice embedding: a one-to-one homomorphism which preserves logical conjunctions and disjunctions, but not negations. The agent's initial probabilities are "extended" to the new algebra, a process which determines what the old belief state has to say about the wider set of possibilities the agent now confronts. This process of extension takes place without considering any new evidence that the agent learns during the experience that brings about her awareness growth. After her initial probabilities have been extended to the new algebra, they can be updated in a belief revision process to reflect any information she has learned about the new possibilities, or the relations between new and old possibilities.

2 Reverse Bayesianism

The approach just outlined is a development of a recent line of investigation, beginning with economists Karni and Vierø, and culminating in Richard Bradley's recent book *Decision Theory with a Human Face*. My development is motivated in part by recent criticisms of this approach by Steele and Stefánsson and Mahtani. I will therefore briefly introduce Karni and Vierø and Bradley's original proposals.

Karni and Vierø's is a complex, choice-based treatment of growing awareness, but it's epistemic component comes down to the following proposal.

> **Reverse Bayesianism.** Suppose that A and B are propositions the agent was previously aware of, P represents the agent's prior beliefs, and P^+ represents their extension after a growth of awareness. For any such A, B, where $P(A) > 0$ and $P(B) > 0$, a rational agent will have:
>
> $$\frac{P(A)}{P(B)} = \frac{P^+(A)}{P^+(B)}$$

Reverse Bayesianism (RB) is presented (e.g., by Vallinder (2018) and Steele and Stefánsson) as a form of conservative belief revision: in the face of new information, the

agent responds appropriately to the new information while preserving as much as possible from their prior belief state. In this case, what is preserved is the ratio of probabilities of propositions they were previously aware of.

In some ways RB is a weak constraint. It tells you only how to constrain the ratios of probabilities of propositions you were previously aware of. It says nothing about the probabilities of new propositions. Another way of stating this is that the permissible posterior belief state, according to RB, is highly imprecise: it is a set of probability functions, each of which obeys RB, but which differ on the new propositions. This is sensible: an agent who was previously unaware of certain possibilities has no real way of constraining their attitudes toward those possibilities.[2]

Here is an example of the principle in action.

> *Weather*. Naledi is considering tomorrow's weather. Being South African, she is aware of three possible kinds of weather: rain, clouds, sun. But having just moved to Sweden, she becomes aware of a fourth kind of weather: snow. Her awareness state grows from the three propositions represented by {RAIN, CLOUDS, SUN} to the four propositions {RAIN, CLOUDS, SUN, SNOW}[3]

RB demands that, for example, $P^+(\text{HOT})/P^+(\text{COLD}) = P(\text{HOT})/P(\text{COLD})$, but says nothing about $P^+(\text{SNOW})$.

Richard Bradley's proposal is similar. Bradley argues that the conservatism we see in three Bayesian belief revision rules (Bayes, Jeffrey and Adams updating) involves the rigidity of conditional beliefs. So, he concludes, extending probabilities to a wider algebra should also preserve conditional probabilities. He provides a condition which does this and which vindicates certain intuitions about awareness growth in simple cases. The condition is that "the agent's new conditional probabilities, given the old domain, for any members of the old domain should equal her old unconditional probabilities for these members" (Bradley, 2017, p. 258). Or, to use terminology Bradley introduces, the new belief states must be *rigid extensions* of the old.

Here's a fuller definition of Bradley's proposal:

> [**Rigid Extension.**] Let $\Omega = \langle \mathcal{X}, \models \rangle$ be a Boolean algebra of propositions of which the agent is initially aware and let $\bigvee \mathcal{X}$ be its top element. Let P represent be a probability function defined on Ω. Let $\mathcal{E}$ be some set of propositions not contained in $\mathcal{X}$. Let $\mathcal{Y}$ be the closure of $\mathcal{X} \cup \mathcal{E}$ under the Boolean operations and $\Xi = \langle \mathcal{Y}, \models \rangle$, be a Boolean algebra of prospects based on $\mathcal{Y}$. Note that $\bigvee \mathcal{X}$ belongs to $\mathcal{Y}$. Then, for any P, a corresponding P^+ on Ξ is called a *rigid extension* of P to Ξ iff, for all $X \in \mathcal{X}$,

$$P^+(X | \bigvee \mathcal{X}) = P(X)$$

---

[2]At the moment, this may sound more polemical to precise Bayesian ears than I intend it to be. What I mean is: there's nothing in the *prior* that will help fix the probabilities of the new possibilities. If you're committed to determining unique credences using the principle of indifference, have at it. As will become clear later, there is no conflict with how I interpret Reverse Bayesianism.

[3]Although awareness states are represented by Boolean algebras, when discussing toy examples I'll often refer to the partition of maximally specific propositions that the agent is aware of *as* their awareness states, metonymically.

Bradley takes it to be a norm of rationality that our beliefs are rigidly extended when our awareness grows.

Consider Weather again. Let $\mathcal{X}$ ={RAIN, CLOUDS, SUN}$\wedge${HOT, COLD} be Naledi's initial awareness state. Then if P^+ is a rigid extension of her prior P, it will have $P^+(\text{RAIN}|\bigvee \mathcal{X}) = P(\text{RAIN})$, but again there are no constraints on $P^+(\text{SNOW})$.[4]

While not equivalent to Reverse Bayesianism, Bradley's rigid extension (RE) principle does imply it, on one reading.[5] This has led some authors, such as Katie Steele, Orri Stefánsson, and Anna Mahtani, to *call* Bradley's position Reverse Bayesianism. In any case, counterexamples to RB will invalidate RE.

RE is, roughly, the proposal that I think is correct for awareness growth. But as it stands, its formulation is unclear and underspecified, as the following section will demonstrate.

The key insight in the RE proposal is that awareness growth is (almost) the inverse problem to learning. In a learning problem, an agent excludes possibilities by learning which element of a partition is true. By conditioning on the learned proposition, the agent restricts the support of their credence function to a subset of the initial algebra: the algebra "below" their evidence—the logically strongest proposition that they have credence 1 in.[6]

In an awareness growth case, the agent includes new possibilities and shifts to a wider algebra. Their initial beliefs, defined on the smaller initial algebra, provide limited constraints on this new set of possibilities. In the absence of new *evidence* (i.e., when all that has been introduced is new possibilities), the principle of rationality guiding their extension is this: if the agent later discovers that the new possibilities are not the case, and restricts their beliefs to the set of propositions corresponding to their prior awareness state, they should recover their initial beliefs. (This is "reverse Bayesianism" in the sense that extension should allow some hypothetical subsequent Bayesian learning to recover your current state.) Why? Because they have no new information which could motivate a change in the contents of their beliefs. All they have gained is an awareness of possibilities, not any evidence regarding those possibilities.

3   Awareness Growth

In the literature on awareness growth there is much informal talk about "growing" the agent's algebra, and about the new algebra "containing the old propositions". In this

---

[4]Everything I argue for in this paper is compatible with an imprecise probability representation of belief. Indeed, that is the account I favour and it is the context in which Bradley introduces his Rigid Extension principle. I work in terms of precise priors here for simplicity only.

[5]Here's a misleading "proof". Consider the ratio $P^+(A|\bigvee \mathcal{X})/P^+(B|\bigvee \mathcal{X})$, for $A, B \in \mathcal{X}$ with $P(A) > 0, P(B) > 0$. Note that $A \models \bigvee \mathcal{X}$ and $B \models \bigvee \mathcal{X}$, Thus $P^+(A \wedge \bigvee \mathcal{X}) = P^+(A)$ and $P^+(A \wedge \bigvee \mathcal{X}) = P^+(A)$. So $P^+(A|\bigvee \mathcal{X})/P^+(B|\bigvee \mathcal{X}) = P^+(A)/P^+(B) = P(A)/P(B)$. My later arguments will show why this reasoning is faulty, but I think that it is assumed by the authors I am responding to.

[6]This is one way of thinking of it, in any case—see for example Miklos Redei's work on conditional expectation. If we want to use the algebra to model awareness, we may need to revisit this way of thinking. An agent who learns by conditioning doesn't lose their awareness of the possibilities that they can now exclude. Similarly, they're aware of all of the entailments of their evidence but these sit "above" the top element of the sub-algebra I describe. These issues don't matter for my main argument here, but they are worth investigating.

section I will highlight various problems with this informal discussion, and how it has been treated in formal models.

The background analogy underlying such talk seems to be something like a set of everyday objects. You start off with two apples, and describe them as a set of two apples. You place a third next to them, and note that now you have a set of three. Physically, the original two are still there, and there's one new apple next to them. Similarly, we might say that the set of two apples "grew" and the set of three apples contains the original two.

For cases like Weather, there's some initial plausibility to this: it seems natural to say that Naledi's partition of precipitation propositions went from {RAIN, CLOUDS, SUN} to {RAIN, CLOUDS, SUN, SNOW}—the old propositions, RAIN and so on, are right there next to the new one, SNOW. This kind of case is called an *expansion* in the literature: a partition is expanded by adding a new proposition.

Another archetypical case is distinguished in the literature: *refinement*. Here, a new partition is introduced and the agent comes to make finer distinctions than they did before.

> *Weather 2*. Naledi now has a four element precipitation partition, as before. But she realises that she needs to consider temperature too. She distinguishes two temperatures: hot and cold. Her awareness state grows from the four propositions {RAIN, CLOUDS, SUN, SNOW}, to the eight propositions represented by {RAIN, CLOUDS, SUN, SNOW}∧{HOT, COLD}.[7]

Naledi's awareness has been refined because, where she previously distinguished one possibility RAIN, she now distinguishes two, RAIN∧HOT and RAIN∧COLD.

But now we can see that the simple notion of propositional preservation across awareness changes will not do. For one thing, cases like Weather 2 immediately illustrate the importance of certain modelling choices that are usually unimportant.

Suppose that propositions are thought of as sets of possible worlds. In Weather 2, we initially need four worlds, which we can label in terms of the propositions Naledi is initially aware of: {RAIN, CLOUDS, SUN}. When the temperature partition is added, this completely changes the world-structure needed to model Naledi's awareness. Whereas RAIN was previously a singleton, consisting of a RAIN-world, it becomes a set of two worlds: {RAIN∧HOT, RAIN∧COLD}. The RAIN proposition has different content after the awareness growth. Talking about it as being "the same", and even applying the label "HOT", may be misleading.[8]

Alternatively, we might take propositions to be fundamental and use "worlds" as merely a colourful description for maximally specific propositions. On this way of thinking, we can say "nothing has changed about the propositions RAIN—it remains unchanged by the awareness growth—all that has happened is that we've introduced a new partition and so there are more maximally specific propositions than there used to be, as these are just elements of the joint partition."

But even here there are complications. Consider the proposition RAIN, and its negation ¬RAIN. Before her awareness grew for the first time, Naledi took ¬RAIN to be CLOUDS∨ SUN. But after her awareness grows, ¬RAIN is CLOUDS∨SUN∨ SNOW. We're in a strange position: we can say that the RAIN proposition is "still there" in the new algebra, but its negation has changed. On one way of specifying the proposition, as

---

[7]This notation is to be read as follows: for two sets $\mathcal{X}, \mathcal{Y}$, $\mathcal{X} \wedge \mathcal{Y} = \{x \wedge y : x \in \mathcal{X}, y \in \mathcal{Y}\}$.

[8]Thanks to Chloé de Canson, who first pointed this out in conversation.

¬RAIN, we can say it is "still there"—RAIN of course has a negation. But it is equivalent to different propositions in the old algebra and the new algebra.

We may want to say that the issue here has to do with negation (indeed, this is what I will say later). But we also have to worry about the interaction between how we label and structure propositions in our model, and facts about the agent. In some contexts, it is perfectly legitimate to relabel propositions for convenience, and to talk about whichever partitions are useful for our purposes. Is that the case here? If so, life is more complicated. Suppose that instead of speaking in terms of the partition {RAIN, CLOUDS, SUN} we shift our attention to {DRY, ¬DRY}, where "DRY" is a label used to track all of the events that don't involve rain. So initially ¬DRY is a relabelling of RAIN, while DRY is a equivalent to CLOUDS∨SUN. Now when Naledi becomes aware of the possibility of snow, she distinguishes it from rain and so SNOW ends up in DRY alongside CLOUDS and SUN. So, after Naledi's awareness grows, it is the "plain" proposition DRY which has changed, while its negation ¬DRY remains the same.

Note that these difficulties have nothing to do with belief, or belief revision—there's no mention of P in the few paragraphs above. They concern how to describe changes of awareness at the level of the propositions and algebra involved, and in particular how to identify propositions across such changes. These issues do, however, generate problems for belief revision procedures such as Reverse Bayesianism because the authors who proposed them have not yet resolved these prior questions satisfactorily.

3.1 Mahtani's "splitting" proposition cases

In a recent paper, Anna Mahtani (2020) presents a problem for RB and RE that turns on these concerns. She introduces two awareness growth cases which, looked at one way, involve expansion, while looked at another way, involve refinement. As a result, RB appears to generate two conflicting demands for preserving probability ratios, in a way that disallows assigning any credence to the new possibilities. Here is the first case.

> *The Other Tenant.* Suppose that you are staying at Bob's flat which he shares with his landlord. You know that Bob is a tenant, and that there is only one landlord, and that this landlord also lives in the flat. In the morning you hear singing coming from the shower room, and you try to work out from the sounds who the singer could be. At this point you have two relevant propositions that you consider possible…with LANDLORD standing for the possibility that the landlord is the singer, and BOB standing for the possibility that Bob is the singer.
>
> Because you know that Bob is a tenant in the flat, you also have a credence in the proposition (TENANT) that the singer is a tenant. Your credence in TENANT is the same as your credence in BOB, for given your state of awareness these two propositions are equivalent. Let us suppose, just for simplicity, that your credence in LANDLORD is 0.5 and your credence in TENANT (and so of course in BOB) is 0.5.
>
> Now let's suppose that the possibility suddenly occurs to you that there might be another tenant living in the same flat, and that perhaps that is the person singing in the shower. Let's assume that no other possibilities occur

to you—e.g. it does not occur to you that it might be a visitor singing in the shower, or just a recording, or anything like that.

Mahtani says: "we have a refinement relative to the possibilities LANDLORD and TENANT, but an expansion relative to the possibilities LANDLORD and BOB." (p.7)

As "TENANT and BOB are possibilities that you were aware of before the awareness growth," Mahtani considers applying RB to the TENANT-BOB pair in which case they have the same credence and OTHER gets none. She then considers applying RB to the LANDLORD-BOB and LANDLORD-TENANT pairs, but then again RB will leave no credence for OTHER. Finding this result unreasonable, Mathani takes is as evidence that Reverse Bayesianism cannot be the correct norm for awareness growth.[9]

Before the awareness growth, the alternative to LANDLORD is a single proposition which I will label TENANT/BOB. After the awareness growth there are two alternatives to LANDLORD: BOB and OTHER. Reverse Bayesianism says that the probability ratios between familiar propositions must be preserved. In so doing, it assumes that all of the old propositions are in the new algebra, and that there's no ambiguity about which of the propositions in the new algebra they are. But in this example, it appears that a single proposition from the old algebra (TENANT/BOB) has been replaced by two propositions in the new algebra (TENANT and BOB), in a hierarchical relationship: TENANT has two sub-possibilities, BOB and OTHER. This allows one to find multiple pairs of propositions to feed into the RB recipe, leading to the incorrect prescription that OTHER must receive no credence.

This isn't quite how Mahtani put it, of course. At times, she talks about TENANT and BOB as two names for the same proposition (before the awareness growth). But at other times, she says that you were aware of two propositions before the change: TENANT and BOB. It is just that, "given your state of awareness these two propositions are equivalent." While I find it difficult to state clearly, I think I understand what this means. Before the awareness change, you can distinguish between the claim that Bob is singing and the claim that the tenant is singing (their utterances use different words, they have different senses), and perhaps you recognise that for some speakers they won't be substitutable. Their "equivalence" means that, given what you're aware of, you take them to have the same referent and your attitude to them is the same. I think one could argue that they are the same proposition, but I don't want to go too deeply into a discussion of the nature of propositions.

I also do not wish to disparage the import of Mahtani's challenge to Reverse Bayesianism. I agree that we should expect any account of awareness growth to be able to handle this sort of case, and that RB and RE cannot do so at present. I diagnose RB's problem with these cases differently, however. I think RB is underspecified because it doesn't properly deal with how the algebra itself has changed. In its current form, it admits of at least two specifications. On Mahtani's specification, when an old proposition "splits" into two in the new algebra, the RB rule must be applied to the two split propositions (which means they get equal credence) and to all the pairs these split propositions form with other old propositions. This gets the wrong prescription. On another specification,

---

[9]Mahtani says that Reverse Bayesianism "effectively rules out awareness growth" in this case. I dispute this language use: it seems perfectly possible that one becomes aware of a possibility and simultaneously learns that it is not the case. Awareness is modelled by the proposition's presence in the algebra, not by the probability it is assigned.

we must choose one, and only one, of TENANT and BOB to be "the same as" TEN-ANT/BOB and apply RB accordingly. This won't generate the incorrect prescription, but the problem is how to specify which new proposition to identify with the old.

Mahtani considers a proposal by Steele and Stefánsson for restricting RB to pairs of propositions X, Y such that the new possibilities are evidentially irrelevant to the matter of X versus Y. The introduction of OTHER is irrelevant to LANDLORD versus BOB, on their proposal. But as OTHER is relevant to TENANT, the pairs involving TENANT aren't subject to RB. This gets the intuitively correct result, that the relative probability of LANDLORD and BOB stays fixed, and OTHER is assigned any value of probability you like within that constraint.[10]

But Mahtani introduces another example that they cannot handle, a case with the same structure but a different intuition about which ratio should be preserved. In The Other Tails, an agent considers a coin and wonders whether it will land HEADS or TAILS. They then think about what image is on the tails side of the coin. Initially they think all coins show a lion, so that TAILS and LION are equivalent (as TENANT and BOB were taken to be above). Later, the agent comes to consider the possibility that the tails side has an image of Stonehenge. So in the new awareness state, TAILS has two sub-possibilities, LION and STONEHENGE. The revised Reverse Bayesian prescription is that you must keep your credence in HEADS and LION in the same proportion as before. But here, intuitively, we want the probability ratio of HEADS-TAILS to be preserved—after all, what the image might be is irrelevant to the probability of a coin toss. But HEADS-TAILS is like LANDLORD-TENANT, and in The Other Tenant we concluded that we wanted the LANDLORD-BOB ratio to be preserved. We need an explanation of how our intuitions are settling on which pair's ratio should be preserved, and our rule for belief revision needs to track that explanation.

Again I think that what this shows is that Reverse Bayesianism, and the account that underlies it, is under-specified. There is something right about it—in each case, there is a probability ratio we want preserved. But it doesn't tell us which one that is, nor is it accompanied by a compelling story of why it should be that ratio. I will argue that a fuller analysis of the awareness change, as distinct from the accompanying belief revision, will help to clarify matters and allow us to preserve what is good about RB.

3.2   A model of awareness growth (but not belief revision)

What this shows us is that talking about awareness growth, and specifically about changes to the algebra, is difficult and confusing when done in this informal manner. We need a precise way of defining what it means for a proposition from the old algebra to be "in the new algebra", or put another way, of defining what it means for a proposition from the old algebra to be identical with a proposition in the new algebra.

In order to do that, I will now introduce some mathematical machinery. The agent's awareness state is modelled by a Boolean algebra, which philosophers are used to thinking of as a kind of field of sets. I will instead consider them algebraically, by which I mean that I will be thinking of a Boolean algebra as a complemented distributive lattice of propositions, ordered by an implication relation.

---

[10]I'll come to Steele and Stefánsson in more detail later on, but for the moment want to continue drawing out the implications of Mahtani's examples.

Here is a brief introduction to lattice theory terminology. Considered algebraically, a lattice is a mathematical structure $\langle \mathcal{X}, \wedge, \vee \rangle$, consisting of a set $\mathcal{X}$ and two operations, called meet ($\wedge$) and join ($\vee$). Meet and join are associative and commutative, each is idempotent, and they obey an absorption law.[11] A distributive lattice is one where the meet and join operations distribute over one another. A bounded lattice has distinguished elements, $\perp$ and $\top$, here called bottom and top respectively, such that $X \wedge \perp = \perp$, and $X \vee \top = \top$ for all $X \in \mathcal{X}$. A complemented distributive lattice is a bounded lattice such that, for each $X \in \mathcal{X}$, there is a unique element of $\mathcal{X}$, denoted $\neg X$, such that $X \wedge \neg X = \perp$ and $X \vee \neg X = \top$. Note that, while the distinguished elements are typically denoted $\top$ and $\perp$, in the context of multiple algebras it is useful to make the algebra explicit and thus to denote them $\bigvee \mathcal{X}$ and $\bigwedge \mathcal{X}$ respectively.

In our case, $\mathcal{X}$ will be a set of propositions. The notation above is deliberately suggestive: meet and join are equivalent to the logical operations of conjunction and disjunction. In some contexts I will talk in terms of an implication relation, $\models$, defined by $X \models Y$ iff $X \wedge Y = X$ iff $X \vee Y = Y$. $\models$ is also called the "order" for the lattice.[12]

Lattices can be visualised in Hasse diagrams. A Hasse diagram represents each element of the lattice, and draws a line that goes upward from X to Y whenever $X \models Y$. There is no interpretation of lines crossing, all the matters is the start- and end-point of any line. Three simple lattices are shown in Figure 1. Note some differences between the three examples. In 1(a), $P \vee Q = \top$, so if we take this to be a *complemented* lattice, $\neg P = Q$. In 1(b), by contrast, $P \vee Q \neq \top$, and instead $\neg P = Q \vee R$. In 1(c), I've shown a case where P and Q are independent propositions, each with their own complement labelled in the usual way.

So, Boolean algebras are particularly rich examples of lattices. As is well known, this richness is crucial to defining probabilities. In the probabilistic belief models I am interested in, Boolean algebras represent awareness states: each element of the algebra is a proposition of which the agent is aware. In cases of awareness change, both the initial and final awareness states will be represented by Boolean algebras. The question this section addresses is: what is the relation between the initial and final algebras? This is the reason that I am interested in lattice theory, which mostly considers sets equipped with less structure than full Boolean algebras: because awareness *changes* only partially preserve the Boolean algebra structure.

Recall the Weather example above, where after Naledi's awareness change, the negated proposition ¬RAIN had changed. This is because in some kinds of awareness growth (often called *expansion* cases), the agent comes to realise that what they previously took

---

[11] Idempotency means that $X \vee X = X$, and $X \wedge X = X$. The absorption laws are $X \vee (X \wedge Y) = X$, $X \wedge (X \vee Y) = X$.

[12] Some may be more familiar with the order-theoretic perspective (with which there is no conflict). I prefer to highlight the algebraic properties of lattices because it makes the link with propositional logic clearer, and I find it fits more naturally with concerns about morphisms, to be introduced below.

Nonetheless, here is the order-theoretic definition of a lattice. A lattice is a poset in which every pair of elements has a meet and a join. A poset is a set X equipped with a partial ordering $\lesssim$. A partial ordering on a set $\mathcal{X}$ is a reflexive, anti-symmetric and transitive binary relation $\lesssim$ on the elements of $\mathcal{X}$. If $X \lesssim Y$ then we say X is less than or below Y, and Y is greater than or above X.

The meet of $X, Y \in \mathcal{X}$, $X \wedge Y$, is the greatest lower bound of X and Y: the greatest element of $\mathcal{X}$ lying below both X and Y. The join of $X, Y \in \mathcal{X}$, $X \vee Y$, is the least upper bound of X and Y: the least element of $\mathcal{X}$ lying above both X and Y.

Note that, as the partial order is reflexive, $X \wedge X = X = X \vee X$ for all $X \in \mathcal{X}$. It is also easy to see that $X \lesssim Y$ iff $X \wedge Y = X$ iff $X \vee Y = Y$.

Figure 1: Hasse diagrams showing three simple lattices. (a) Two non-trivial elements in a single, two-element partition. (b) Three atoms in a single, three-element partition. (c) The Boolean algebra cl($\{P, Q\}$).
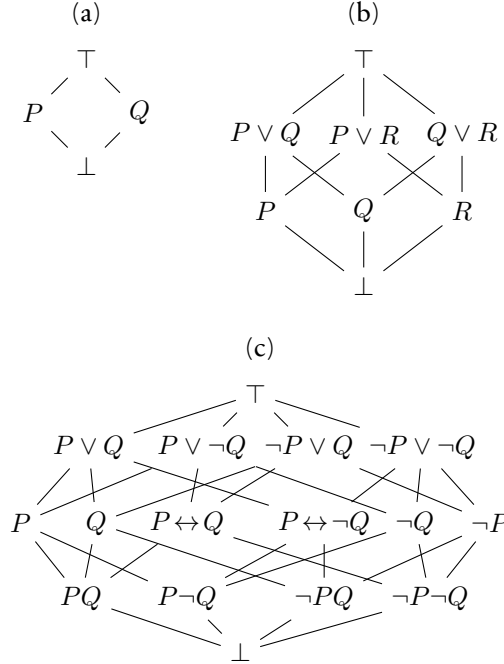
to be the total set of possibilities is incomplete. Naledi realises that {RAIN, CLOUDS, SUN} isn't a complete set of precipitation states—and so it isn't certain that RAIN$\lor$CLOUDS$\lor$SUN. While RAIN$\lor$CLOUDS$\lor$SUN was the top element of the algebra representing Naledi's initial awareness state, it is not the top element of the algebra that represents her new awareness state.

We want to relate the old algebra to the new, in a way that tracks our intuitions about which propositions are "still there" while also tracking this change of complementation structure. For this we need a mapping between the two. As we're considering awareness growth, where new possibilities are added, what we need is a mapping called a *lattice embedding*.

> **Lattice embedding.** A map $h : \mathcal{L} \to \mathcal{K}$, between two lattices $\langle \mathcal{L}, \lor, \land \rangle$ and $\langle \mathcal{K}, \lor, \land \rangle$, is a lattice embedding iff it is one-to-one lattice homomorphism. That is, a one-to-one map that is meet- and join-preserving: $\forall X, Y \in \mathcal{L}, h(X \lor Y) = h(X) \lor h(Y), h(X \land Y) = h(X) \land h(Y)$.

A lattice embedding maps each proposition from the old algebra to a proposition in the new algebra, and which preserves the lattice operations, meet and join.[13] Above I've made their respective meet and join operations explicit, and put a bar on the operations in the codomain algebra. The fact that that it is one-to-one means that the image $h(\mathcal{L})$ is a sublattice of $\mathcal{K}$ which is isomorphic to $\mathcal{L}$.[14] Note that a lattice embedding does *not* preserve complementation, i.e., there is no guarantee that $h(\neg X) = \neg h(X)$.

---

[13]Relating to the order-theoretic definition of lattices: defined this way, lattice homomorphisms are order-preserving (Davey and Priestley, 2002, Proposition 2.19, p.44)

[14]A sublattice of a lattice $\mathcal{L}$ is a subset $\varnothing \neq \mathcal{M} \subseteq \mathcal{L}$, such that $\mathcal{M}$ is closed under the lattice operations:

In Weather, Naledi's initial precipitation partition is {RAIN, CLOUDS, SUN}, and her final precipitation partition is {Rain, Clouds, Sun, Snow}. (In order to make the point I want to here, I've changed the capitalisation of the final versions.) The natural embedding maps h(RAIN)=Rain, h(CLOUDS)=Clouds and h(SUN)=Sun. The fact that h is a lattice homomorphism guarantees that h(RAIN ∨ CLOUDS) = Rain ∨ Clouds and h(RAIN ∧ SUN) = Rain ∧ Sun. In this latter case, the conjunction of RAIN and SUN is a contradiction, RAIN ∧ SUN = ⊥, and our mapping gets us this too as, in the new algebra, Rain ∧ Sun = ⊥. However, in the initial algebra ¬ RAIN = CLOUDS ∨ SUN, whereas in the new algebra ¬ Rain = Clouds ∨ Sun ∨ Snow. So h(¬ RAIN) = h(CLOUDS ∨ SUN) = h(CLOUDS) ∨ h(SUN) = Clouds ∨ Sun ≠ ¬ h(RAIN). As advertised, the lattice embedding does not preserve complementation which tracks the result of our informal model and our intuition about what has changed for Naledi.

We've now arrived at our recipe for modelling awareness growth.

> **Modelling recipe.** An agent's awareness state is modelled by a Boolean algebra $\Omega = \langle \mathcal{X}, \vee, \wedge \rangle$. After their awareness grows, they have a new awareness state: the Boolean algebra Ξ. We relate the old algebra to the new by embedding Ω into Ξ. The one-to-one association of propositions in Ω with propositions in Ξ ensures that the old propositions are "in" the new algebra: there is an embedding h such that, for each $X \in \Omega$, $h(X) = x \in \Xi$. (Ξ must of course also contain the new propositions that are the content of the awareness change.)

The recipe does not tell us, at this stage, which embedding we need.

As this model will do significant work in criticising previous authors writing on awareness growth, I want emphasise why we should represent things this way. First, when comparing algebraic structures and their elements, the natural notion of "identity" is isomorphism. The lattice embedding is an isomorphism between the old algebra and its image in the new algebra. It is the right kind of isomorphism, because it preserves meet and join but not complementation. Second, I have highlighted difficulties with the informal, intuitive notion of identity between propositions in different algebras, so *some* new account is needed. I claim that mine has various benefits. It gives a clear answer to what it means to identify a proposition from the old algebra with a proposition in the new algebra. It also explains what is puzzling about Mahtani's cases, and provides a resolution.

3.3 Mahtani's cases, again

Let's start by thinking about Mahtani's statement that there are three propositions in your initial awareness state in The Other Tenant: LANDLORD, TENANT, and BOB. In a lattice theoretic presentation we can't represent BOB and TENANT as two different propositions, at best they're two labels for the same proposition. That's because propositions are elements of the set $\mathcal{X}$ that is the basis of the algebra Ω, and sets contain only one copy of each item. Or, from another direction: because Ω is a Boolean algebra, each element must have a *unique* complement—if TENANT and BOB have the same complement (LANDLORD) then they must be identical.

$\forall X, Y \in \mathcal{M}, X \vee Y \in \mathcal{M}, X \wedge Y \in \mathcal{M}.$

I don't mean to hide behind the formalism here; this isn't meant to be a "computer says no" kind of response. The whole theory of probability is built on Boolean algebras, and these just aren't the kind of structures that can represent the kind of distinct-but-equivalent propositions that Mahtani gestures at. I don't think that is a problem either. As I will show, we can capture the sense in which the original "TENANT" proposition "splits" when your awareness grows.

The definition of identity comes directly from the lattice homomorphism. Consider again Ω and Ξ, with h a lattice embedding between them. A proposition $Y \in \Xi$ is "the same as" a proposition $X \in \Omega$ just in case h maps X to Y: $h(X) = Y$. As h is one-to-one, every proposition in Ω is associated with one in Ξ, and as it is a homomorphism, the disjunction and conjunction relations between propositions are preserved.

What is puzzling about Mahtani's cases is that they offer us two choices for how to embed the original algebra into the new algebra. But *this problem has nothing to do with Reverse Bayesianism.*

In Figure 2(a), I've shown the initial algebra Ω with the two possibilities you start off being aware of: L for LANDLORD and TB for TENANT/BOB. In this model, there is simply no way to represent Mahtani's description of the initial awareness state as containing a proposition BOB and another TENANT, which coincide in some sense. The best we can do is to recognise the two ways of labelling the one proposition, which I've done with the composite label TB. Figure 2(b) and (c) show the new algebra Ξ, in which there's a composite possibility, TENANT (t), with two sub-possibilities, BOB (b) and OTHER (o). I've used lower-case labels for these propositions in Ξ, because I want to insist that they're mathematically different entities, which will come to be associated with the elements of Ω only via an embedding.

Here is one possible embedding $h : \Omega \to \Xi$, $h(L) = l, h(TB) = b$. Under this embedding, the old possibility TENANT/BOB is mapped to the new possibility BOB. We can work out the rest from the fact that h is a lattice homomorphism. So $h(\perp) = h(L \wedge TB) = h(L) \wedge h(TB) = l \wedge b = \perp'$. Importantly, $h(\top) = h(L \vee TB) = h(L) \vee h(TB) = l \vee b$, which is not the top element of Ξ! That is, of course, what we want: you previously thought that the Landlord and Bob were the only possibilities for the singer, but then you learned that there is another possibility. This also changes the complementation structure of the algebra: now you realise that, if it is not the Landlord singing, then it might be Bob or the Other Tenant. Previously you thought that if it were not the Landlord, it must be Bob. The image of Ω in Ξ, under this embedding, is shown in Figure 2(b) in blue. This embedding makes The Other Tenant an example of expansion.

Another embedding h' maps TENANT/BOB to the new possibility TENANT. It is shown in Figure 2(c) in red. On this embedding, The Other Tenant is a case of refinement: where you previously thought in terms of "the tenant" you now recognise two finer distinctions within this proposition.

Although this section is just about awareness changes and "growing an algebra", it will help see the purpose of all this is we take a quick detour into belief revision. (Mahtani, after all, is arguing against Reverse Bayesianism.) First, note that the RB principle itself needs to be adjusted. It is framed in terms of "the same proposition" appearing in the old and new algebra, which we now see is just a shorthand; convenient but potentially misleading. Here is a revised principle, tracking the relevant lattice embedding:

> **Reverse Bayesianism (new).** Let Ω be an agent's initial algebra, and Ξ their final algebra, related by a lattice embedding $h : \Omega \to \Xi$. Let P be the agent's

Figure 2: The Other Tenant. (a) The old algebra. (b) The new algebra, with the old embedded into the blue portion. (c) The new algebra, with a different, red, embedding. On Mahtani's analysis, (b) is the preferred embedding for The Other Tenant.
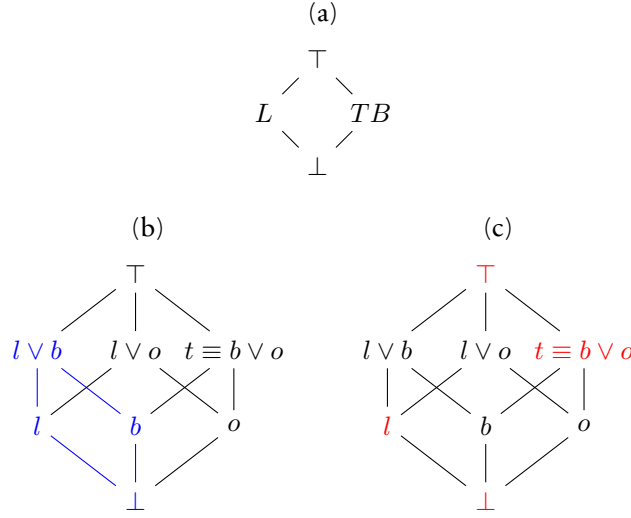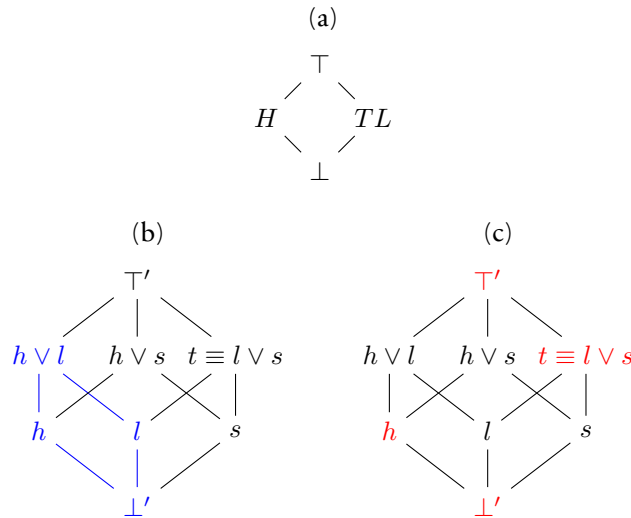

Figure 3: The Other Tails. (a) The old algebra. (b) The new algebra, with the old embedded into the blue portion. (c) The new algebra, with a different, red, embedding. On Mahtani's analysis, (c) is the preferred embedding for The Other Tails.

prior credence function on Ω, and let P^+ be a probability function on Ξ. For any $A, B \in \Omega$, where $P(A) > 0$ and $P(B) > 0$, P^+ is a rational extension of P to Ξ iff:

$$\frac{P(A)}{P(B)} = \frac{P^+(h(A))}{P^+(h(B))}$$

If we apply Reverse Bayesianism (new) to the two embeddings I introduced above for The Other Tenant, we get different results. On h (blue) we get:

$$\frac{P(L)}{P(TB)} = \frac{P^+(h(L))}{P^+(h(TB))} = \frac{P^+(l)}{P^+(b)}$$

Whereas on h' (red) we we get:

$$\frac{P(L)}{P(TB)} = \frac{P^+(h'(L))}{P^+(h'(TB))} = \frac{P^+(l)}{P^+(b \vee o)}$$

Notice that we don't ever get the situation Mahtani discusses: where we are forced to assign TENANT and BOB the same probability by RB. That's because embeddings are one-to-one mappings, so there is no embedding which will map TENANT/BOB to TENANT and to BOB. The sense in which TENANT/BOB "is" the proposition TENANT *and* "is" the proposition BOB is just that there exists an embedding on which it is mapped to each of them. So our revised RB principle won't problematically assign the new possibility zero credence, in the way that Mahtani highlighted.

We can now evaluate these two embedding options. (We would like a systematic way of doing this, but for now I'll keep it intuitive and informal.) The h embedding requires that there is no change in the relative probabilities assigned to the singer being the Landlord or Bob. So, whatever credence is assigned to the possibility that there is another tenant who is the singer, it needs to get its probability mass equally from that previously assigned to LANDLORD and TENANT/BOB. Recalling that Mahtani has the priors set up with $P(TB) = P(L) = 0.5$, this means that the extended probability LANDLORD and BOB must be identical: $P^+(l) = P^+(b) = 0.5 - k$, $P^+(o) = 2k$.

The h' embedding, on the other hand, requires that the extended probability for LANDLORD and TENANT be identical. TENANT has two sub-possibilities, BOB and OTHER. So where you previously assigned equal credence to the singer being the Landlord or the tenant (who you took to be Bob), now that you are aware of the possibility of another tenant, you assign equal credence to the singer being the landlord or *either* tenant: $P^+(l) = P^+(t) = 0.5$ as before. There are no constraints on the credence assigned to b and o so long as $P^+(b) + P^+(o) = 0.5$.

The latter assignment seems, to me, to be the less sensible. Mahtani agrees: "given that there might be two tenants, it is natural to suppose that your credence in TENANT should increase relative to LANDLORD" (p.9). Incidentally, it is also the result given by Steele and Stefánsson's restriction of RB.

In The Other Tails, we face a similar choice. Recall that in this example, you previously had two labels, TAILS/LION, for one of your propositions. When your awareness grows, you end up with a proposition TAILS that has two sub-possibilities, LION and STONEHENGE. We can therefore embed your old algebra into the new one using a (suitably defined) h embedding, shown in blue in Figure 3(b), or with an h' embedding shown in red in Figure 3(c). In this case, however, our intuitions are that h' is the more sensible embedding: as I said above, the number of potential images on the tails side doesn't change the probability of the coin toss result.

3.4 Choosing an embedding

While the structure that I have provided above clarifies the problem of awareness growth significantly, we would still like guidance on which embedding is the right one for a particular example. This would be valuable philosophically, and is of practical necessity for any who would wish to translate my eventual proposal into algorithms for machine learning purposes. Unfortunately I am not sure this kind of exercise can be solved in perfect generality. What I can offer is some guidance, based on an examination of the cases above and the reasoning that seems to underlie the intuitions about the correct embeddings.

I begin with The Other Coin. This example has a particular feature that the others do not: the agent's subjective probabilities for the coin toss are derived from the chances we commonly take to hold for coin tosses. These chances plausibly derive from the symmetry of the coin itself: its two sides and constant weight distribution. They therefore "attach" to the {HEADS, TAILS} way of specifying the possibilities. When the agent extends their beliefs to the new algebra, they ought to do so in a way which respects the chance structure that determined their original credences. So this is a case in which we are choosing an embedding—which is a way of specifying how the agent's *awareness* has grown—based on belief considerations. In particular we look at the reasons that their beliefs are structured as they are, and choose an embedding which allows us to carry these reasons over to the new algebra.

Now let us consider The Other Tenant. Here, Mahtani tells us that the agent starts by assigning equal probability to LANDLORD and TENANT/BOB. If we take the lesson from the coin case to apply here, we should look for the reasons that underlie this assignment of credence and see how they apply to our choice of embedding. Often such assignments of equal credence are due to reasoning according to a "principle of indifference", on which one assigns equal probability to each outcome. Can this allow us to distinguish between the {LANDLORD, BOB} partition and the {LANDLORD, TENANT} partition? Perhaps. Applications of the principle of indifference require specifications of "outcomes" which are considered equivalent. On the {LANDLORD, BOB} way of thinking, the reasoning is presumably that there are two people in the house, that there is no reason to suppose one of them is more likely to be showering at the moment, and therefore to assign them equal probability. On the {LANDLORD, TENANT} way of thinking, the reasoning would have to involve the two roles in the legal agreement governing Bob's occupation of this apartment. It seems much less plausible that the agent's initial reasoning attached itself to the legal roles of "landlord" and "tenant" than to the two individuals qua people. Therefore, when we consider which embedding to choose on the basis of the reasons that fixed the agent's prior beliefs, the embedding which identifies propositions on the basis of personhood is preferred to the embedding which identifies propositions on the basis of legal role.

This, then, is my proposal about embedding selection: we choose the embedding which best preserves the reasoning underlying the agent's initial credal assignments.

4 Belief Revision

In this section I will show how the model above, along with the multi-stage analysis of awareness growth experiences that it fits naturally with, delivers the right answer to a set

of cases that Steele and Stefánsson (forthcoming) have recently introduced as challenged for the Rigid Extension principle.

4.1   Steele and Stefánsson's evidential relevance cases

Steele and Stefánsson claim that RE is violated by "examples where awareness grows since an agent becomes aware of a proposition that she takes to be evidentially relevant, intuitively speaking, to the comparison of propositions of which she was already aware" (p.55). They offer a few examples of this sort, but I will focus on the one I take to be the most challenging.

> *Relativity*. Nineteenth century physicists were unaware of the Special Theory of Relativity [SR]. That is, not only did they not take the theory to be true; they had not even entertained the theory. We can suppose, however, that they had entertained various propositions for which the theory was regarded evidentially relevant, once Einstein brought the theory to their attention. In particular, they did (rightly) take the theory to be evidentially relevant to various propositions about the speed of light, such as whether the speed of light would always be measured at 300,000 km/s independently of how fast the investigator is moving or whether the measured speed would differ, depending on how fast the investigator is moving. But then the awareness and subsequent acceptance of [SR] changed their relative confidence in such propositions. (p.55-56)

As Steele and Stefánsson interpret the example, the agent understands that SR is evidentially relevant to the proposition "light's speed is constant", LC. The thought is that, after their awareness grows to include SR, we don't want $P^+(LC)/P^+(\neg LC) = P(LC)/P(\neg LC)$, but rather we expect that $P^+(LC)/P^+(\neg LC) > P(LC)/P(\neg LC)$, as SR is a theory on which light's speed is absolute, and so any allocation of probability mass to SR will make that LC more likely than it was before. They conclude "that we should not impose Reverse Bayesianism as a general constraint on how a rational agent can revise her credences when her awareness grows" (p.57).

Let us note the nature of the example. Steele and Stefánsson speak of the new proposition (SR) being "evidentially relevant" to an old one (LC), but in this case the link is stronger. The constancy of the speed of light is a postulate of Special Relativity, and so SR logically entails LC.

I will argue that there are two problems with Steele and Stefánsson's analysis of the example. Firstly, they fail to properly appreciate the requirements for identifying old propositions with new, as outlined in section 3. Secondly, and independently, their analysis only holds if one insists that RE cover all the changes that need to take place following awareness growth. In other words, if one assumes that these cases must be analysed in one go, rather than in stages as I described above, and if one assumes that RE is a proposal for such a single-stage analysis.

Let us begin with the familiar. The way in which Steele and Stefánsson take Relativity to be a counter-example to Bradley's Rigid Extension reveals why we need the embedding framework I introduced above. Recall that Rigid Extension requires that the agent's new conditional probabilities, given the top element of the old algebra, for any members of

Figure 4: Recreating of Steele and Stefánsson's diagram showing the awareness growth in Relativity, as they see it.
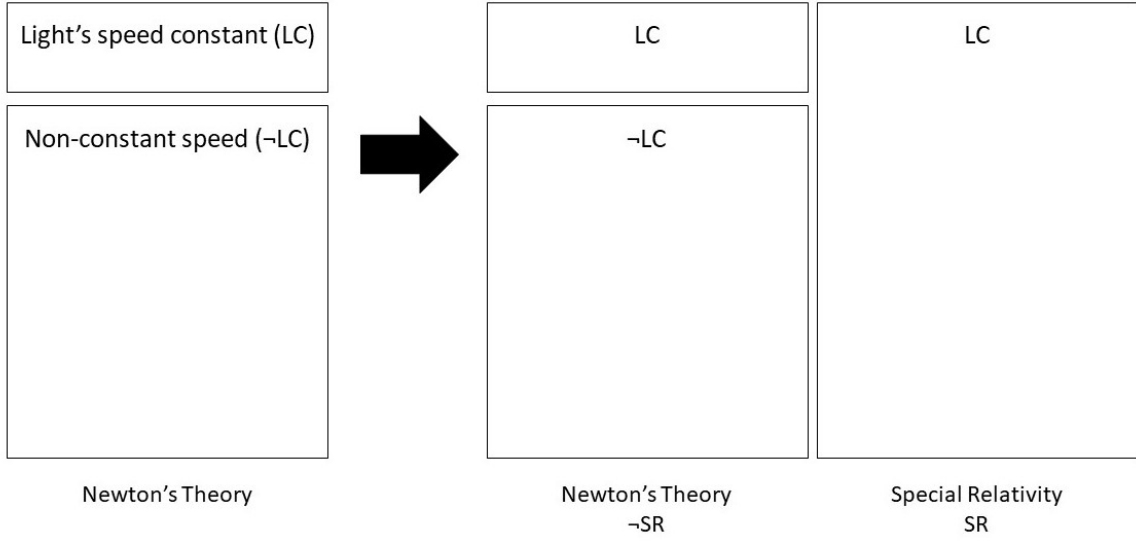
the old algebra, should equal her old unconditional probabilities for these members. Here is an updated version of this principle, reflecting the lattice embedding:

> **Rigid Extension (new).** Let $\Omega = \langle \mathcal{X}, \models \rangle$ be the agent's initial awareness state, and P be their prior credence function. Let $\Xi = \langle \mathcal{Y}, \models \rangle$ be the new algebra after their awareness grows, and P^+ be their extended credences on Ξ. Then P^+ is called a *rigid extension* of P iff, for all $X \in \mathcal{X}$,

$$P^+(h(X)|h(\vee \mathcal{X})) = P(X)$$

Steele and Stefánsson in drawing out some differences between a version of RE and Reverse Bayesianism, write: "[Figure 4] makes vivid that [Rigid Extension] does not generally entail Reverse Bayesianism, at least for cases of awareness growth by expansion. [RE] requires that the probabilities for old propositions—[LC] and [¬LC]—*conditional on 'Newton's theory'*, remain constant when awareness grows. One can see just by looking at the figure that it does not follow that the ratio of the absolute probabilities for [LC] and [¬LC] remain constant when awareness grows" (p.64).

On my embedding view, Steele and Stefánsson's application of RE to the Relativity case is incoherent. In order to associate the old propositions with the new, we need to specify an embedding to take us from the initial awareness state to the new one. But their description of the problem corresponds to no possible embedding. To see why, consider Figure 5: 5(a) shows the initial awareness state, in which there is only one theory and two possibilities concerning the speed of light (lc and $\neg lc$). As in their diagram, Newton's theory is the only option and therefore is equivalent to the tautology. In Figure 5(b) we have the new algebra, in which there are two theories (SR for Special Relativity, and ¬SR for Newton's theory), and two propositions concerning light (here, LC and ¬LC). The task before us is to embed the algebra shown on the left into the algebra on the right.

In the quote above, Steele and Stefánsson implicitly use an embedding on which the top element of the initial algebra ($\top$) is mapped to "Newton's theory" (¬SR) in the

refined algebra *and* which maps "light is constant" (lc) to the full LC proposition in the new algebra. No homomorphism can accomplish this. Any embedding h on which $lc \mapsto$ LC will preserve the will map $\top \mapsto \top'$, because it needs to preserve the join structure:

$$h(l \vee \neg l) = h(l) \vee h(\neg l) = \text{LC} \vee \neg\text{LC} = \top'$$

It is possible to map $\top \mapsto \neg\text{SR}$, but then "light is constant" must be mapped to the proposition beneath $\neg$SR: $lc \mapsto \neg\text{SR} \wedge \text{LC}$. Indeed, this seems sensible: before becoming aware of Special Relativity the scientist took Newton's theory to be the only option, but after their awareness grows they see it is merely one of two possible accounts: $\top \mapsto \neg\text{SR}$. But on this way of thinking, you have to give up the identity between the old "light is constant" proposition and the new, higher up, proposition denoted LC. You can't have it both ways.

Figure 5: Hasse diagrams for the initial algebra for Relativity, and the refined algebra after awareness growth.
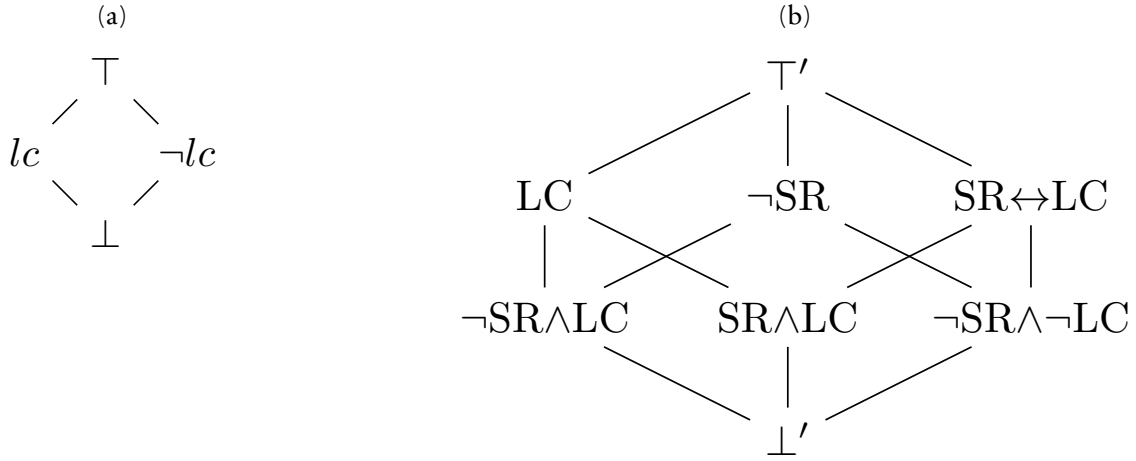
Now let us address the application of Rigid Extension to the case. This case doesn't illustrate the point quite in the way Steele and Stefánsson want, because LC is a postulate of Special Relativity. So in this case, the "evidential relevance" is handled entirely by the embedding process. What it means for a logically omniscient agent to "become aware" of SR and LC is to fit them into an algebra such as that shown in Figure 5(b). But, as embeddings are one-to-one homomorphisms, we don't ever end up in the situation Steele and Stefánsson consider problematic.

But in other examples this won't occur, when the relevance of a new proposition to an old is merely evidential rather than logical. Consider the following case.

> *Movies.* Suppose you are deciding whether to see a movie at your local cinema. You know that on the day in question, the cinema only shows "international" (non-English) movies. You realise that both the movie's language and genre will affect your viewing experience. The possible languages you consider are French and German and the genres you consider are thriller and comedy. But then you realise that, due to your poor French and German skills, your enjoyment of the movie will also depend on the level of difficulty

of the language. Since you know the owner of the cinema to be simple-minded, you are, after this realisation, much more confident that the movie will have low-level language than high-level language. Moreover, since you associate low-level language with thrillers, this makes you more confident than you were before that the movie on offer is a thriller as opposed to a comedy. (Steele and Stefánsson, forthcoming, p. 58)

Steele and Stefánsson write: "The important feature of the above example is that the original awareness context is partitioned according to some property (the language level) that is taken to be evidentially relevant to the comparison of some pair of inconsistent basic propositions—that the movie is a thriller and that it is a comedy—in the old awareness context" (Steele and Stefánsson, forthcoming, p. 58). RE will say that, in such cases, the ratio of probability between Thriller and Comedy ought to remain constant. But, argue Steele and Stefánsson, this should not be—instead we should end up with the credal assignment represented schematically in Figure 6.

I begin again with a comment on the nature of the case. This is a case of forgotten or neglected information, in which the agent not only has previous acquaintance with the "new" propositions, but also to evidential relations between "new" and old propositions. Steele and Stefánsson propose to model it *as if* it were awareness growth, which I agree is possible. But to do so we must treat the "new" information as if it were really new. We therefore proceed as follows: we identify the initial and final algebras, we find an embedding to map the old algebra into the new, we extend the agent's probabilities by RE, and then we apply the new information that is "learned" (in this case, remembered) in a belief revision stage.

Here the "learned" information has two parts. Firstly, "you associate low-level language with thrillers", which we can model as coming to have Q(Thriller | Low)=1. (Note that in this case we assume this is a contingent empirical fact.) Secondly, "you know the owner of the cinema to be simple-minded, you are, after this realisation, much more confident that the movie will have low-level language than high-level language": Q(Low) > Q(High). Each is a piece of information that "comes to you" in your remembering.

Figure 6: Steele and Stefánsson's diagram showing the awareness growth in Movies, as they see it.
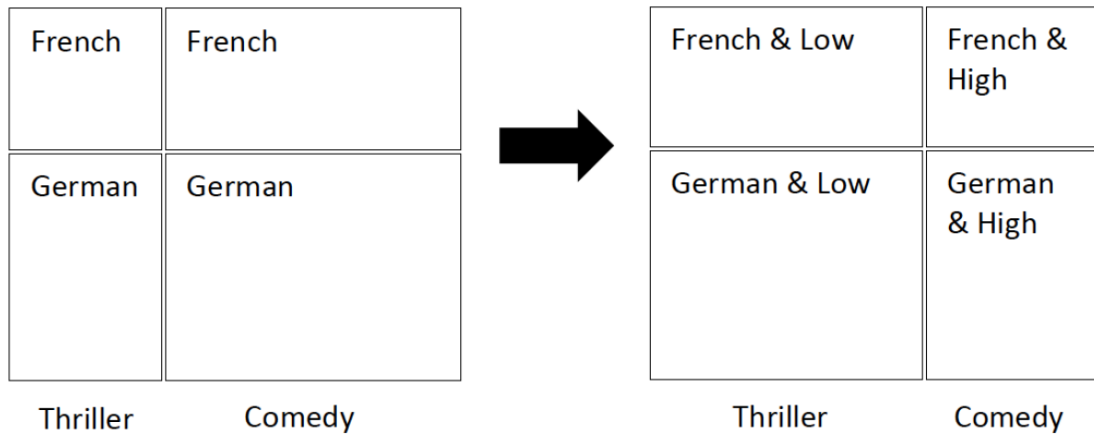
I do not deny that Figure 6 shows the intuitively correct result given all the information presented in Movies. However, I propose that RE should be regarded as only a partial specification of the change to the agent's belief state. Specifically, it deals with the change due *only* to the change of awareness. Put another way, it is a rule for *extending* belief following awareness growth, not for revising it in light of new evidence. Belief revision will be handled separately.

On this view, RE says nothing about how you should change your belief upon learning that a new possibility is evidentially relevant to an old possibility. So it shouldn't be expected to hold *after* new things are learned. It is neither a ban on future learning, nor a one-stop-shop for belief revision in the face of growing awareness.

Semi-formally (i.e., without employing my embedding model) here is how I would apply the rule in this case. The awareness growth takes the agent from an awareness state {French, German}∧{Thriller, Comedy} to final awareness state {French, German}∧{Thriller, Comedy}∧{Low, High}. It is at this stage that RE (or indeed RB) applies. It tells us how to extend the agent's probabilities to the new algebra, and entails that we shouldn't change the ratio of probabilities of known propositions. i.e., The probability ratio of Thriller to Comedy should remain constant. But that isn't the end of the story. The agent also comes to have $Q(\text{Thriller} \mid \text{Low}) = 1$ and $Q(\text{Low}) > Q(\text{High})$. Once these changes are incorporated, it is clear that $Q(\text{Thriller}) > Q(\text{Comedy})$, as expected.

There are three stages to the analysis. The first stage involves apprehending the new propositions and recognising their logical relations to known propositions. It is modelled by forming the relevant algebras, and choosing an embedding which associates the known propositions with propositions in the new algebra. With this done, we move to the second stage and extend the prior attitudes to the new awareness state in line with the demands of rationality. (I agree with Bradley that the right demand is RE.) Having done so, we can update the agent's beliefs in light of the new information learned during the experience in the third stag. The separation is conceptual rather than temporal: the aim is to focus first on the purely awareness-related aspects of the experience and then to turn to the attitude of belief. The experience that brings about these three steps might (and often will) be a single, unitary experience.

The purpose of the separation is analytical clarity, allowing us to distinguish between the agent's priors on the old algebra (P), their extension to the new algebra (P^+), and the posterior after learning (Q). This distinction is necessary: the agent can't be updating P with the new information, as it isn't defined on the right algebra.

Steele and Stefánsson are aware of this option. They reject (a two-stage variant of) it on grounds of ad hocness:

> It might be argued that our examples are not illustrative of a simple learning event (a simple growth in awareness); rather, our examples illustrate and should be expressed formally as complex learning experiences, where first there is a growth in awareness, and then there is a further learning event that may be represented, say, as a Jeffrey-style or Adams-style learning event. In this way, one could argue that the awareness-growth aspect of the learning event always satisfies Reverse Bayesianism (the new propositions are in the first instance evidentially irrelevant to the comparison of the old basic propositions). Subsequently, however, there may be a revision of probabilities over some partition of the possibility space, resulting in more dramatic changes to the ratios of probabilities for the old basic propositions. The reason we re-

ject this way of conceiving of the learning events described by our examples
is that the two-part structure is ultimately unmotivated. The second learning
stage is an odd, spontaneous learning event that would be hard to rationalise.
Hence, this would again seem to us to be an artificial and ad hoc way to save
Reverse Bayesianism. (p.59)

"Spontaneity" seems like the wrong protest. There is nothing "spontaneous" about
the learning event—it is brought about by the experience of, for example, becoming
aware of the new scientific theory. It also seems fairly plausible that in many cases one
does become aware of something before learning about its evidential relations to other
matters.[15]

This is one place where my focus on *true* unawareness is important. In their more
prosaic examples like Movies, Steele and Stefánsson draw on background knowledge of
evidential relations between new propositions and old. As an agent comes to consider a
possibility (which they were previously aware of but neglected), they employ this back-
ground knowledge. But these "new" propositions aren't new in the sense I'm interested
in, they were merely forgotten or neglected until this point. If we want to model such
cases as if they were awareness changes, we need to treat the evidential relations as some-
thing learned. Not because they are truly learned anew, but because we're using a model
developed for a different kind of case (true awareness growth) to model prosaic cases
about forgotten, excluded, or neglected possibilities.

Another reason to separate the two stages is that they don't always occur together.
There are cases of awareness that don't involve learning and others (like Movies) that do.
Insisting that the two be treated identically seems, to me, unmotivated.

4.2 A general model of belief revision

In the discussion of Movies above, I brushed over an important question. I claimed that
we need to separate out the stages of awareness growth and learning, but I didn't address
how this learning takes place. How should we model agents who learn about the new
propositions that they become aware of?

We must start by recognising that learning after (or, as part of) awareness growth can-
not take place by Bayesian conditioning. When an agent Bayesian learns E, they update
their beliefs from P to Q, such that $Q(X) = P(X|E)$, while holding their conditional
probabilities fixed $Q(\cdot|X) = P(\cdot|X)$ for all $X \in \Omega$. But in an awareness growth case, E
is a new proposition. Which is to say $E \notin \Omega$ and no conditional probabilities $P(\cdot|E)$ are
defined. This is part of the very setup of awareness growth. So Bayesians who venture
into these waters must give up their familiar comforts.

So, if not Bayesian learning, then what? We need a much more general model of
belief revision, that doesn't assume priors involving the learned proposition and ideally
one which can handle different kinds of learning: recall that in the examples above,
the agent's learning involved fixing a conditional probability linking an old and new
proposition. I will adopt a simple and general picture of belief revision that accomplishes
these goals. It is inspired by how Richard Jeffrey models cases of uncertain learning:
instead of modelling what was learned as a proposition, Jeffrey proposes that we *describe*

---

[15] Again note the problematic nature of the Relativity example in which LC is a postulate of SR. But
imagine for the moment that SR were merely evidentially relevant to LC.

the effects of the experience on the agent, by stipulating their credences over a partition after the experience. He then provides a rule for generating a fully-specified, unique and coherent posterior credence, now called Jeffrey conditioning (Jeffrey, 1983, p. 165).

Following the formalism of Dietrich, List, and Bradley (2016), I will think of a belief revision rule as a function, mapping an initial belief state and an experience to a final belief state. Let $\mathcal{P}$ be a set of possible belief states, $\mathcal{I}$ be a set of possible inputs or experiences, so that a belief revision rule maps $\mathcal{P} \times \mathcal{I} \to \mathcal{P}$. Belief states will be probability functions, or sets of probability functions, as before. As we're interested in belief revision following an extension to a new algebra, we will typically begin with a highly imprecise state. "Inputs" are taken to be very general, including straightforward observations, noisy signals, expert reports of various kinds. We therefore specify them *extensionally*, as the set of belief states that are consistent with the experience. A simple example: if I look out the window and see that it is cloudy, this input constrains my belief state to include only those in which it is cloudy outside my window.

Belief revision rules can be characterised by two conditions: Responsiveness and Conservatism (Dietrich, List, and Bradley, 2016). Loosely, Responsiveness ensures that the final belief state "respects" the input, and Conservatism ensures that the belief revision changes *only* what is "required" by the input. This is captured by a Conservation condition, that specifies which parts of the prior belief state must be conserved by the revision.

Rules which obey these two conditions follow a pattern called perturbation-propagation.[16] First, as the rule is *Responsive* to the experience, the input will directly bring about a change in belief state: the perturbation. Second, the remainder of the belief state is adjusted to reflect the impact of the input; this propagation makes use of the perturbation and the parts of the initial state that are preserved by the *Conservatism* of the revision. Table 1 shows two common examples: Bayesian updating and Jeffrey updating. Note that the propagation step covers what is typically thought of as a "belief revision rule," such as updating by Bayesian conditioning.

Table 1: Two belief revision rules

Rule	Perturbation	Propagation
Bayes	$Q(E) = 1$	$Q(X) = P(X|E)$
Jeffrey	$Q(A) = \pi_A, \forall A \in \mathbb{A}$	$Q(X) = \sum_{A \in \mathbb{A}} P(X|A)Q(A)$

In generalised belief revision theory, a *kind of experience* is matched with a *particular revision rule*. Dietrich, List, and Bradley (2016) characterise the class of Bayesian inputs as those experiences which constrain the agent's belief state to include only probability functions in which the probability of a specific proposition—the one the agent learns during the experience—is 1. We can similarly define the class of Jeffrey inputs, corresponding to Jeffrey updates. In each case, we can model this with a domain $\mathcal{D}$ that is a subset of the space of possible experiences and initial states: $\mathcal{D} \subseteq \mathcal{P} \times \mathcal{I}$.

Responsiveness consists in ensuring that the final belief state is in the set I, i.e., that it is compatible with the experience. Conservatism is harder to spell out. Each domain that Dietrich, List, and Bradley consider comes with a specification of what those experiences are "silent" on. This notion of silence is used to fill out the norm of Conservatism: put

---

[16]I take this term from Bradley (2017).

roughly, a belief revision rule should leave unchanged whatever the experience is silent on. Dietrich, List, and Bradley (2016) then prove characterisation results showing that for Bayes, Jeffrey and Adams updating there is a unique rule respecting Responsiveness and Conservatism and that in each case it is the rule referred to parenthetically above.[17]

In awareness growth cases, we don't want to restrict the kinds of information the agent may receive about the new possibilities, for this would restrict the applicability of the model. Agents may come to know that one event is more likely than another, or that two variables are independent, or they might learn the expected value for some variable. (van Fraassen (1981) says that learning the expected value for a variable is the most general kind of constraint on your probability function, and that others can all be framed as special instances of it.) So I want my theory of belief revision to be able to handle all of these report types.

The problem is that for these more general input domains, no unique belief revision rule is known. Put extensionally, the problem is that once we identify the set of belief states that respect the input, we lack general rules for further refining this set. This may be easier to see by switching to an intensional definition: let us denote the information learned in the experience with a formula, ϕ_Q. Responsiveness tells us that the posterior credence function must respect this constraint: we want a Q that respects ϕ_Q. But a great many functions will typically do this! What we want is one which also fits the extended prior, P^+, in the right way. What is that way? Conservatism is meant to provide the answer: in the way that preserves as much of P^+ as possible while respecting Q.

Specifying the Conservatism norm for a form of experience is a complex matter. The canonical examples mentioned above have a particularly nice form: each comes with a constraint and a *rigidity* condition which realises the Conservation condition. These conditions, summarised in Table 2, are necessary and sufficient conditions for updating according to the associated belief revision rule (Bradley, 2017, pp. 188–200).

Table 2: Conservatism conditions for Bayes and Jeffrey updates. $\mathbb{A}$ is a partition.

	Constraint	Rigidity condition(s)
Bayes	$Q(E) = 1$	$Q(\cdot\|X) = P(\cdot\|X), \forall X \in \Omega$
Jeffrey	$Q(A) = \pi_A, \forall A \in \mathbb{A}$	$Q(\cdot\|A) = P(\cdot\|A), \forall A \in \mathbb{A}$

For general learning (not in cases of awareness growth) the challenge is this: the agent has priors for all propositions in the domain; *some* of these are replaced by ϕ_Q and this process breaks the coherence of the function. In order to restore coherence, we need to "fit" the remaining parts of P with the bits of Q specified by ϕ_Q. But there are many ways of doing this. In a Bayesian learning case the agent's prior $P(E)$ is replaced by $Q(E) = 1$. Making only this change and keeping the rest of P would lead to incoherence. Coherence is restored by the propagation rule: set the rest of Q equal to P conditional on E.

But for more general kinds of inputs (e.g., learning that expectation of a random variable), we don't have Conservation conditions that produce such unique "kinematic" update rules. One option is to to do more work to identify what is conserved by each kind of experience (i.e., to identify Rigidity conditions like those in Table 2) in order

---

[17]Adams updating is a form of learning in which an agent comes to have new conditional probabilities over a partition. It was introduced by Richard Bradley (2005, 2017).

to formulate a kinematic revision rule. Another option is to find the coherent function which obeys ϕ_Q and is closest to P by some measure of "distance". There is a large literature on using divergences between probabilities for epistemological or decision theoretic purposes: such methods are present in foundational work by de Finetti and Savage; van Fraassen (1981) and Diaconis and Zabell (1982) utilise such methods; and they are present throughout the "accuracy" programme in epistemology including in work by Joyce (1998) and Pettigrew (2016).

This is a significant challenge for belief revision theory generally. But things are much simpler in awareness growth cases, where the learning we are interested in concerns the new awareness. That is because the constraint ϕ_Q concerns the *new propositions*. In the examples we have looked at, the agent comes to have certain posterior probabilities $Q(E_i)$ for a set of new propositions $\{E_i\}$, or conditional probabilities $Q(X_j|E_i)$ where the X_j propositions can be new propositions or (embeddings of) old propositions. But the agent had no priors involving the $\{E_i\}$ propositions, that's part of what it means to say that they underwent a growth of awareness. So there is no conflict with any part of P, or P's extension to the wider algebra P^+. Recall that P^+ is typically very imprecise, a set of many probability functions which match the demands of Rigid Extension on the embedding of the old algebra, but which differ on the (rest of the) new algebra. So we can simply further restrict this set by imposing the belief constraint ϕ_Q gained in the learning experience.

Here is our final belief revision recipe:

> **Generalised belief revision.** An agent has undergone an awareness growth, taking them from initial awareness state Ω and belief state P, to new awareness state Ξ and rigidly extended belief state P^+, via an embedding h. They update their beliefs in light of the learning experience ϕ_Q by adopting posterior belief state Q: the set of credence functions on Ξ which coincide with P^+ on $h(\Omega)$ and which obey ϕ_Q elsewhere. (Again, this set will typically not be a singleton.)

Recall the Movies case. Here the "learning" experience had two effects: $Q(\text{Thriller} \mid \text{Low})=1$, and $Q(\text{Low}) > Q(\text{High})$, representing your associations between Thrillers and basic language on the one hand, and your assumptions about the tastes of the cinema owner for simplistic language on the other. So long as we read these as names of propositions in the new algebra, these two constraints constitute ϕ_Q.

Nothing more is required. We form a new algebra, in which an independent language-level partition {Low, High} is introduced. There is a natural embedding, mapping Thriller to Thriller, and so on. The agent's prior is rigidly extended to the new algebra, which yields the preservation of ratios expected by RB. Then the extended beliefs P^+ are updated in line with ϕ_Q. The endpoint is the intuitive result—that shown in Steele and Stefánsson's diagram, Figure 6. But we get there in stages: first by determining the new algebra, then by applying Rigid Extension (new), and finally by carrying out the generalised belief revision procedure.

References

Alchourrón, C.E., P. Gärdenfors, and David Makinson (1985). "On the Logic of Theory Change: Partial Meet Contraction and Revision Functions". In: *Journal of Symbolic Logic* 50, pp. 510–530.

Bradley, Richard (Apr. 2005). "Radical Probabilism and Bayesian Conditioning". In: *Philosophy of Science* 72.2, pp. 342–364. ISSN: 0031-8248, 1539-767X. DOI: 10.1086/432427. URL: https://www.journals.uchicago.edu/doi/10.1086/432427 (visited on 12/04/2018).

— (2017). *Decision Theory with a Human Face*. Cambridge University Press. ISBN: 978-1-107-00321-7.

Davey, B. A. and H. A. Priestley (2002). *Introduction to Lattices and Order*. 2nd ed. Cambridge: Cambridge University Press.

Diaconis, Persi and Sandy L. Zabell (1982). "Updating Subjective Probability". In: *Journal of the American Statistical Association* 77.380, pp. 822–30.

Dietrich, Franz, Christian List, and Richard Bradley (Mar. 1, 2016). "Belief Revision Generalized: A Joint Characterization of Bayes' and Jeffrey's Rules". In: *Journal of Economic Theory* 162, pp. 352–371. ISSN: 0022-0531. DOI: 10.1016/j.jet.2015.11.006. URL: http://www.sciencedirect.com/science/article/pii/S0022053115002008 (visited on 10/04/2019).

Earman, John (1992). *Bayes or Bust: A Critical Examination of Bayesian Confirmation Theory*. Cambridge, MA: MIT Press.

Eels, Ellery (1985). "Problems of Old Evidence". In: *Pacific Philosophical Quarterly* 66, pp. 283–302.

Glymour, Clark N. (1980). *Theory and Evidence*. Princeton: Princeton University Press. URL: https://catalog.princeton.edu/catalog/9336 (visited on 02/01/2019).

Jeffrey, Richard (1983). *The Logic of Decision*. 2nd ed. University of Chicago Press.

Joyce, James M. (1998). "A Nonpragmatic Vindication of Probabilism". In: *Philosophy of Science* 65, pp. 575–603.

Levi, Isaac (1977). "Subjunctives, Dispositions and Chances". In: *Synthese* 34, pp. 423–455.

— (1980). *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance*. MIT Press.

— (1991). *The Fixation of Belief and Its Undoing*. Cambridge, MA: Cambridge University Press.

Mahtani, Anna (2020). "Awareness Growth and Dispositional Attitudes". In: *Synthese*. DOI: 10.1007/s11229-020-02611-5.

Pettigrew, Richard (2016). *Accuracy and the Laws of Credence*. Oxford: Oxford University Press.

Shimony, A (1970). "Scientific Inference". In: *The Nature and Function of Scientific Theories*. Ed. by R Colodny. University of Pittsburgh Press, pp. 79–172.

Steele, Katie and H. Orri Stefánsson (forthcoming). *Beyond Uncertainty*. Cambridge University Press.

Vallinder, Aron (2018). "Bayesian Variations: Essays on the Structure, Object, and Dynamics of Credence". PhD Thesis. London: London School of Economics and Political Science.

Van Fraassen, Bas (Dec. 1, 1981). "A Problem for Relative Information Minimizers in Probability Kinematics". In: *The British Journal for the Philosophy of Science* 32.4, pp. 375–

379. ISSN: 0007-0882. DOI: 10.1093/bjps/32.4.375. URL: https://academic.oup.com/bjps/article/32/4/375/1516900 (visited on 09/11/2019).

Wenmackers, S and Jan-Willem Romeijn (2016). "New Theory About Old Evidence". In: *Synthese* 193.4, pp. 1225–1250.