Awareness growth and belief revision

Joe Roussos

March 2021

Abstract

How should an agent update their beliefs when they encounter a completely new possibility? Bayesianism has no answer, as it insists that agents have priors for all possibilities. This essay develops a two-stage approach to awareness growth, each based on norms of conservative belief change. The first stage is a form of Reverse Bayesianism, a proposal for extending one's current credences to a new set of possibilities. My model resolves a challenge by Mahtani, that Reverse Bayesianism yields the wrong result when a proposition "splits" due to the change in awareness. The second stage involves revising these extended beliefs. Steele and Stefánsson, argued recently that Reverse Bayesianism cannot deal with new propositions that are evidentially relevant to old propositions, but I show that these cases are easily handled in my two-stage model. I conclude with a discussion of the rational grounding of general, non-Bayesian, belief revision in this context.

1 Introduction

Ordinary people like you and me regularly confront new possibilities. When I moved to Stockholm, I learned that Riddarholmen lies just west of Gamla Stan. Climate scientists in the twentieth century explored new theories and mechanisms, such as the runaway greenhouse effect, to explain observed climate phenomena and predict climate change. These each involved the formation of new beliefs. The configuration of Stockholm's islands involves a set of prosaic propositions that I had simply never encountered or considered before. In the climate case, the scientists involved developed and learned entirely new concepts.

It is a strange failing of our most popular formal models of belief that they have little to say about this kind of learning. I speak here of "Bayesian" models of belief, those that represent beliefs with probabilities and insist that learning is accomplished by conditioning. Bayesianism is a rich and successful theory in philosophy, statistics, and more recently machine learning. But in Bayesian models, all resolutions of uncertainty take place by updating pre-existing beliefs. Agents must have priors for propositions to learn about them at later stages. In this way, Bayesianism leaves no room for agents to learn about genuinely new states of affairs and has no guidance for real agents when they undergo such changes of awareness. New possibilities are at once so common and so bound up in our most pressing epistemic challenges that filling this gap is a matter of first importance to epistemology.

I will set out a model of growing awareness which specifies how an agent's probabilistic beliefs ought to be extended to a new space of possibilities, and which provides constraints for the formation of new beliefs about those new possibilities. My focus will be entirely on degree of belief, leaving "full" belief aside (thus "belief" always refers to degree of belief here). I will also say nothing of desire or decision. The model I present is normative, and the guiding normative principle will be conservative belief change.

Our starting point is a probabilistic model of rational belief, which has least two components: an algebra of propositions and a probability function defined on that algebra. The probability function represents the agent's degrees of belief. Awareness, as I will think of it, is a necessary condition for taking an attitude to a proposition.[1] As the probability function represents the attitude (belief), it is a natural suggestion to use the algebra itself to model the agent's state of awareness, which I will sometimes also call their "awareness context". I will say that a proposition within the algebra is one that the agent is aware of, and that they are not aware of any propositions not in the algebra.[2] Changes to awareness will be thus be modelled as changes to the algebra.[3]

[1]It may also be a sufficient condition, but I think there is reason to doubt this. Consider a comparativist framework in which comparative partial beliefs are primary and probabilistic credences mere representations of them. If incomplete partial beliefs are rationally permitted then an agent might be aware of a proposition without their partial belief relation including it. Even though the proposition would be in the domain of an imprecise probabilistic representor of that agent, we might still want to say that the agent does not truly have an attitude to that proposition.

[2]This is no constraint on extending the model to include desire and decision, so long as we adopt a propositional decision theory such as that developed by Richard Jeffrey (1983).

[3]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 et al. (1985).)

Two notes on doing this this way. First, in such models the algebra also encodes the logical relationships between propositions. Though I am interested in agents who are more realistic than the Bayesian ideal, I will preserve the idealisation that they are logically omniscient: their degrees of belief reflect all the deductive logical relationships between propositions of which they are aware. New propositions will be automatically slotted into the appropriate places in a new algebra. I do not aim to model agents who become aware of new logical connections, nor agents who become aware of new possibilities but do not grasp their logical relations to known possibilities.

Second, this setup means that agents have no attitudes to propositions they are unaware of. This is deliberate: I am interested in *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, however, a narrower definition of awareness growth than some have used. In different ways, Bradley (2017), Mahtani (2020), and Steele and Stefánsson (2020, n.d.) all consider cases in which agents do have attitudes to propositions they are "unaware" of. These authors want to cover cases of forgetting, or failure to consider a proposition, or of deliberately excluding a proposition from consideration. This invites confusion, precisely because the phenomena are quite different with respect to the agent's possession of prior attitudes to the "new" possibilities. That said, I acknowledge that, while agents who have 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 ought to recognise the significant conceptual difference between an agent who is truly unaware of a proposition—such as a scientist in the moments just prior to first learning of a new theory—and, say, an agent who is temporarily excluding something from consideration.

This confusion is compounded by the use toy cases involving small sets of prosaic propositions, even when discussing true unawareness. It can be 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 using the intuitions attached to them. I shall pick my way through this thicket by considering only true unawareness, and asking you to reject those holdover intuitions in toy examples.

Our topic is how an agent's beliefs should change when they become aware of new propositions. A variant of this problem has been discussed under the description "the problem of new hypotheses" in the philosophy of science (it is linked to the "problem of old evidence" raised by Glymour (1980)). As the name implies, the focus there was on hypotheses and their confirmation, which limited the scope of the discussion somewhat. Previous work in that context includes that by Shimony (1970), Eels (1985), Earman (1992), and more recently Wenmackers and Romeijn (2016) who adopt a Reverse Bayesian solution. This topic is the probabilistic companion to a much discussed issue in logic and computer science, concerning the logic of unawareness and qualitative belief revision following awareness growth; Schipper (2015) is a thorough review. As my focus is on probabilistic belief, my attention will be on a more recent literature in decision theory and epistemology.

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 awareness is again represented by such an algebra. Propositions from the old algebra are identified with propositions in 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. In most cases it is obvious which propositions are preserved across awareness changes, and when more care is needed I propose that the reasons underlying the agent's credal structure fix this identification. 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. As the agent has no priors for the new propositions, this belief revision cannot be Bayesian. I show that that the same logic of conservative belief change which underlies ordinary Bayesian conditioning and the Reverse Bayesian belief extension procedure can guide belief revision following awareness growth.

In section 2, I introduce Reverse Bayesianism, my favoured approach to the extension step described above. Section 3 considers awareness growth (but not belief revision), and responds to Anna Mahtani's recent challenge to Reverse Bayesianism. Section 4 takes up belief revision; responding to Katie Steele and Orri Stefánsson, and developing a general framework for belief revision after awareness growth.

2 Reverse Bayesianism

Consider this case:

> *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}[4]

Suppose that, for Naledi, RAIN, CLOUDS and SUN were mutually exclusive and exhaustive weather possibilities to begin with. Assume further that she was genuinely not aware of the possibility of SNOW. Finally, let us assume that like a good Bayesian Naledi has priors for each of these possibilties; perhaps she takes them to be equally likely so that $P(\text{RAIN}) = P(\text{CLOUDS}) = P(\text{SUN})$.

[4]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.

When Naledi becomes aware of the possibility of SNOW, she will come to hold some belief about it. The question we are interested in is this: what constraint does rationality put on the assignment of credence to SNOW, in light of Naledi's prior beliefs?

The approach I defend falls under the heading "Reverse Bayesianism". This is a cluster of approaches, first developed by the economists Karni and Vierø (2013, 2015, 2017). Reverse Bayesian approaches have been developed in philosophy by Wenmackers and Romeijn (2016) and Bradley (2017), and applied by Vallinder (2018). I will follow Bradley's development, and will simply refer to the approach I outline below as Reverse Bayesianism, without trying to track the differences and disputes amongst Reverse Bayesians.

Reverse Bayesianism (RB) is presented (e.g., by Vallinder (2018) and Steele and Stefánsson (2020, n.d.)) as a form of conservative belief revision: in the face of a stimulus, the agent responds appropriately to the stimulus while preserving as much as possible from their prior belief state. The signature feature of RB is the requirement that the agent preserves the ratios of probabilities of propositions they were previously aware of. So for Naledi, the 1:1 ratios between RAIN, SUN and CLOUDS should be preserved. Learning of the mere possibility of SNOW should not change how relatively plausible the known weather states are. In some ways this is a weak constraint. It says nothing direct about the probabilities assigned to the new possibilities. Another way of saying this is that the permissible extended 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 reasonable way of constraining their attitudes toward them.[5]

This preservation of ratios is merely a sign of Reverse Bayesianism, however. Here is how I will define it:

> **Reverse Bayesianism:** When an agent undergoes awareness growth, their belief state should be rigidly extended to the new possibility space.

[5]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 are 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.

A *rigid extension* is defined by Bradley as follows: "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, 258). Here is a first pass at a formal statement of this. Let $\mathcal{X}$ be the set of all propositions of which the agent is initially aware, and $\bigvee \mathcal{X}$ be its top element, the tautological disjunction of the agent's possibilities. Let P be the probability function on $\mathcal{X}$ representing their credences.[6] The agent becomes aware of some new propositions contained in set $\mathcal{E}$. We form a new set $\mathcal{Y}$, the closure of $\mathcal{X} \cup \mathcal{E}$ under the Boolean operations. Note that $\bigvee \mathcal{X}$ is in $\mathcal{Y}$. We now consider probabilities defined on $\mathcal{Y}$.

> **Rigid Extension:** For any P, a corresponding P^+ on $\mathcal{Y}$ is called a *rigid extension* of P to $\mathcal{Y}$ iff, for all $X \in \mathcal{X}$, $P^+(X|\bigvee \mathcal{X}) = P(X)$.

Consider Weather again. Let $\mathcal{X}$ ={RAIN, CLOUDS, SUN} be Naledi's initial awareness state. Then if P^+ is a rigid extension of her prior P, it will have $P^+(\text{RAIN}|\bigvee \mathcal{X}) = P(\text{RAIN})$, and the same for SUN and CLOUDS. Note again that there are no constraints on $P^+(\text{SNOW})$.[7]

Why think Reverse Bayesianism is the right constraint for rational awareness growth? My aim is not to provide a ground-up defence of the principle here, so I will only briefly review the justification I find most convincing: RB follows from the logic of conservative belief change that underlies other norms for rational belief revision. [8]

In the face of new information, a rational agent updates their beliefs to accommodate what has been learned but does no more; going beyond the demands of the evi-

[6]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.

[7]While not equivalent to the preservation of ratios that others take to *define* Reverse Bayesianism, Bradley's definition does imply it under certain conditions. 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.

[8]Another notable justification is Karni and Vierø's axiomatisation of Reverse Bayesianism. But theirs is a complex, choice-theoretic presentation and would take us too far afield. Suffice to say that it involves strong assumptions, including that the agent has a precise probability and utility function after the awareness change. Agents which meet this and other rationality constraints are required to obey the ratio preservation formula, which is a consequence of Reverse Bayesianism as I define it here (Karni and Vierø, 2013).

dence is unjustified. This prescription of minimal change in the face of new information is sufficient to generate Bayesian conditioning and Jeffrey conditioning as special cases, appropriate to the information learned in each case (Williams, 1980). In cases of awareness growth, no proposition has been learned. Instead, the agent has become aware of a proposition or possibility. So how can we apply this logic of conservatism here?

Bradley provides the necessary argument. He considers a case like *Weather*, and argues that an agent's prior credences reflect the relative plausibility of the prospects they are aware of. These judgements imply specific degrees of belief, but these cannot be compared across different sets of propositions. The form of belief change that Naledi undergoes when she becomes aware of the possibility of snow is not one that requires any change in her relational attitudes: she has learned nothing to imply that her prior judgement about the relative plausibility of rain compared to sunny weather (Bradley, 2017, 255). Bradley then notes that the conservatism we see in three Bayesian belief revision rules (Bayes, Jeffrey and Adams updating) involves the rigidity of conditional beliefs. Conditional probabilities are thus the vehicle for accomplishing the required conservation of prior relational beliefs(Bradley, 2017, 258). But, as there are no prior attitudes to the new possibilities, it is the conditional probabilities of known propositions which are subject to the constraint. Rigid extension is the most demanding conservative requirement available.

A similar argument can be applied to cases of "refinement", in which a new partition refines previously known possibilities. Consider this case:

> *Weather 2*. Naledi now has a four element precipitation partition, as above. 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}.[9]

Naledi's awareness has been refined because, where she previously distinguished one

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

possibility RAIN, she now distinguishes two, RAIN$\wedge$HOT and RAIN$\wedge$COLD. But, argues Bradley, considering different ways one might fine-grain RAIN should not change its possibility; these are just *varieties* of RAIN. The mere fact that one's view of the world was too coarse should not prompt a change in the coarse-grained beliefs. (Bradley, 2017, 257). Reverse Bayesianism delivers this verdict. As Bradley considers any awareness growth to be decomposable into combinations of these two kinds, expansion and refinement, this is a complete argument.[10]

3 Awareness Growth

When an agent's awareness grows, it has become common to talk of the algebra which represents their awareness state as also "growing". Similarly, because awareness growth adds new possibilities, authors writing on this topic talk about the "old propositions" being in the new algebra. In this section I will highlight some difficulties that arise from making this intuitive informal talk precise.

Underlying such talk is an appealing analogy with sets 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. We might say that the original set "grew", and the new set contains the original. There is nothing mysterious here, as apples have clear identity conditions and sets can include one another.

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}. This simple representation is taken to be faithful to what is really happening, and so we take seriously that the old propositions, RAIN and so on, are right there next to the new one, SNOW. But there are two complications, arising from the fact that propositions stand in logical relations, and that these logical relations are what distinguish a simple set from a structured algebra.

[10] Critics agree with Bradley on this point, see e.g., Steele and Stefánsson (n.d.) and Mahtani (2020).

Consider the proposition that it is not raining, denoted ¬RAIN. Before her awareness grew in *Weather*, Naledi took ¬RAIN to be equivalent to CLOUDS∨ SUN. But one way of describing her awareness change is that Naledi learns a new way for it not to rain, so that ¬RAIN is equivalent to CLOUDS∨SUN∨ SNOW. What to say about which propositions are "still there" after the awareness change? On one way of specifying the proposition, as ¬RAIN, we can say it is "still there"—the agent is still aware of the possibility that it is not raining. But what the conditions under which this is true have changed. So we cannot identify propositions on the basis of truth-conditional equivalence across awareness changes. In particular, we will need to treat negations carefully.

Now consider Naledi's change of awareness in *Weather 2*. In epistemological settings, Bayesian models often represent propositions with sets of possible worlds. If we did things this way then we would initially need four worlds, which we can label in terms of the propositions Naledi is initially aware of in *Weather 2*: {RAIN, CLOUDS, SUN, SNOW}. When the temperature partition is added, this completely changes the world-structure needed to model Naledi's awareness. Whereas the proposition RAIN was previously a singleton, consisting of a RAIN-world, it "becomes" a set of two worlds: {RAIN∧HOT, RAIN∧COLD}. What has happened to the RAIN-world? From the third-personal perpective we might say that it has split in two. The RAIN proposition, we could add, has different content after the awareness growth.[11] Thus talking about a proposition as being "the same" across awareness changes may be misleading (as has been pointed out by Steele and Stefánsson, n.d.).

Doing away with "worlds" as primitive objects can reduce this complexity somewhat. On this way of thinking, propositions aren't sets (they may have other internal structure, but I do not model it here), and so RAIN is an object which can be "still there" after Naledi's awareness is refined. The introduction of a new partition means that we no longer take RAIN to be a maximally specific description of ways things might be, but this was an inessential feature to begin with. This is the approach I shall take.

[11] An alternative approach, of starting with all the worlds we need to model the final state, is fraught. It is difficult to rationalise the agent's initial state if we, the modeller, know that their partitions are incomplete.

Note that the complexities discussed here have nothing to do with belief, or belief revision. 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 extension procedures such as Reverse Bayesianism because the authors who proposed RB had not resolved these prior questions satisfactorily.

3.1 Mahtani's "splitting" proposition cases

In a recent paper, Anna Mahtani (2020) presents a problem for Reverse Bayesianism 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 [OTHER]. 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, 2020, 5–6)

How should your credences be redistributed? The trouble comes from TENANT and BOB. As these are propositions you were aware of previously, RB applies to them. As they were equivalent and thus had identical probabilities, RB demands that their probabilities remain identical. After your awareness grows, you recognise that OTHER entails TENANT and is disjoint from BOB. But then OTHER can have no credence assigned to it at all! Reverse Bayesianism is doing too much: forbidding the assignment of credence to a new possibility.[12]

As far as the definition of RB is concerned, TENANT and BOB should be treated as identical—though clearly there is a sense in which they're different. Similarly, TENANT is an "old proposition" and so RB applies to it—though clearly there is a sense in which there is something new about it. These qualifications provide room for a defender of RB to quibble about how Mahtani applies the rule here, but they would be hard pressed to give a principled reason for applying it in such a way as to avoid the bad result. What is needed is a resolution of these questions of propositional identity within an awareness state and across awareness states. But these questions aren't really about Reverse Bayesianism—any account of belief change following awareness growth is going to need an answer to them. Mahtani's challenge is an important one, and she is correct that RB cannot handle these cases at present. But once these questions have been resolved, a minor modification will allow us to preserve what is good about RB.

[12] 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.

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

We will now need some additional mathematical machinery, which will bring with it some helpful precision.

An awareness state will be modelled by a Boolean algebra. Philosophers are used to thinking of as a kind of field of sets. In that framework, the primitive objects are worlds, propositions are sets of worlds, and a Boolean algebra is a set of propositions closed under complementation and union. I will instead consider Boolean algebras as algebraic structures. By this I mean that I will take propositions as primitive objects, and regard a Boolean algebra as a complemented distributive lattice of propositions, ordered by an implication relation.

Here is a brief introduction to this lattice theoretic terminology. Considered algebraically, a lattice is a mathematical structure $\langle \mathcal{X}, \wedge, \vee \rangle$, consisting of a set $\mathcal{X}$ and two operations on it, called meet ($\wedge$) and join ($\vee$). Meet and join are associative and commutative, each is idempotent, and they obey an absorption law.[13]In our case, the objects in $\mathcal{X}$ are propositions. The notation is deliberately suggestive: meet and join are equivalent to the logical operations of conjunction and disjunction.

A *distributive* lattice is one where the meet and join operations distribute over one another, as the logical operations do. A *bounded* lattice has two special elements, denoted $\bot$ and $\top$ and here called bottom and top respectively, defined by $X \wedge \bot = \bot$, and $X \vee \top = \top$ for all $X \in \mathcal{X}$. These correspond to the contradiction and tautology. 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 $X \wedge \neg X = \bot$ and $X \vee \neg X = \top$. This completes our correspondence with propositional logic, by introducing negation.

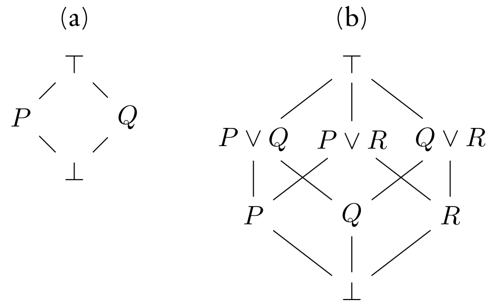
One notational note: in the context of multiple algebras the $\top$ and $\bot$ notation is ambiguous. To make the algebra they belong to explicit, I will sometimes denote them $\bigvee \mathcal{X}$ and $\bigwedge \mathcal{X}$ respectively. I will also sometimes talk in terms of an implication relation, $\models$, defined by $X \models Y$ iff $X \wedge Y = X$ iff $X \vee Y = Y$. $\models$ is also called the "order" for the

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

lattice.[14]

Lattices can be visualised in Hasse diagrams. A Hasse diagram has a node for each element of the lattice, and draws a line that goes upward from X to Y whenever $X \models Y$. All the matters is the start- and end-point of any line, there is no interpretation of lines crossing. Two simple lattices are shown in Figure 1. Note one differences between them: 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 $\neg P = Q \vee R$.

Figure 1: Hasse diagrams showing two simple lattices. (a) Two non-trivial elements in a single, two-element partition. (b) Three atoms in a single, three-element partition.



Boolean algebras are particularly rich examples of lattices, and this richness is crucial to defining probabilities on them. Now recall the interpretation I am using: each element of a Boolean algebra is a proposition of which the agent is aware. In cases of awareness change, both the initial and final awareness state will be represented by a Boolean algebra. The question this section addresses is: what is the relation between the initial and final algebras? Recall the *Weather* example above, where after Naledi's awareness change, the operation ¬RAIN picks out a different partition than it did before. This is because in expansion cases the agent realises that what they took to be the total set of

[14]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 $X \vee Y = Y$.

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$\vee$ CLOUDS$\vee$SUN. While RAIN$\vee$CLOUDS$\vee$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. This is the reason that I am interested in lattice theory, which mostly considers sets equipped with less structure than full Boolean algebras. When awareness changes, some structure is preserved but not all of it. In particular, the complementation relations can change, as can the identity of the bounds.

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 algebras. As we're considering awareness growth, where new possibilities are added, we need an injective map. In order to be an identity criterion for propositions, it needs to be one-to-one. And as some structure is preserved, it is natural to look at the class of lattice homomorphisms. Putting this together, we are led to consider *lattice embeddings*.

> **Lattice embedding.** A map $h : \mathcal{L} \to \mathcal{K}$, between two lattices $\langle \mathcal{L}, \vee, \wedge \rangle$ and $\langle \mathcal{K}, \bar{\vee}, \bar{\wedge} \rangle$, 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: $\forall X, Y \in \mathcal{L}, h(X \vee Y) = h(X) \bar{\vee} h(Y), h(X \wedge Y) = h(X) \bar{\wedge} h(Y)$.

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.[15] Above I've made their respective meet and join operations explicit, and put a bar on the operations in the codomain algebra. I won't be as careful with the notation from here on, because using an embedding means that there's no need to differentiate between $\vee$ and $v\bar{e}e$. Because an embedding is one-to-one, the image $h(\mathcal{L})$ is a sublattice of $\mathcal{K}$ which is isomorphic to $\mathcal{L}$.[16] Note that a lattice embedding does *not* preserve complementation, i.e., there is no

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

guarantee that $h(\neg X) = \bar{\neg} h(X)$.

Let's apply this new machinery. In *Weather*, Naledi's initial precipitation partition is {RAIN, CLOUDS, SUN}, and her final precipitation partition is {Rain, Clouds, Sun, Snow}. (Note the new capitalisation.) The natural embedding maps

$$\text{RAIN} \overset{h}{\mapsto} \text{Rain}$$
$$\text{CLOUDS} \mapsto \text{Clouds}$$
$$\text{SUN} \mapsto \text{Sun}$$

The fact that h is a lattice homomorphism guarantees that h(RAIN $\vee$ CLOUDS) = Rain $\vee$ Clouds and h(RAIN $\wedge$ SUN) = Rain $\wedge$ Sun. In this latter case, the conjunction of RAIN and SUN is a contradiction, RAIN $\wedge$ SUN = $\bot$, and our mapping gets us this too as, in the new algebra, Rain $\wedge$ Sun = $\bot$. However, in the initial algebra $\neg$RAIN = CLOUDS $\vee$ SUN, whereas in the new algebra $\bar{\neg}$ Rain = Clouds $\vee$ Sun $\vee$ Snow. So $h(\neg$ RAIN) = h(CLOUDS $\vee$ SUN) = h(CLOUDS) $\vee$ h(SUN) = Clouds $\vee$ Sun $\neq \bar{\neg}h$(RAIN). This tracks the result of our informal model, and more importantly 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 $\Omega = \langle \mathcal{X}, \vee, \wedge, \neg \rangle$. After their awareness grows, they have a new awareness state: the Boolean algebra $\Xi = \langle \mathcal{Y}, \vee, \wedge, \bar{\neg} \rangle$. 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 $X \in \Omega, h(X) = x \in \Xi$. ($\Xi$ 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.

[16]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: $\forall X, Y \in \mathcal{M}, X \vee Y \in \mathcal{M}, X \wedge Y \in \mathcal{M}$.

3.3 Mahtani's cases, again

I will now apply this recipe to Mahtani's case of *The Other Tenant*. Recall that we start off with an initial awareness state consisting of LANDLORD, TENANT, and BOB. In a lattice theoretic presentation we simply cannot represent BOB and TENANT as two different propositions if we are to capture the other things Mahtani says about them: that they are equivalent and that they therefore carry the same probability. At best we can say that these are two labels for the same proposition. There are two ways to see that we're forced into this choice: first, 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. Second, Ω is a Boolean algebra and so each element must have a *unique* complement—if TENANT and BOB have the same complement (LANDLORD) then they must be identical.

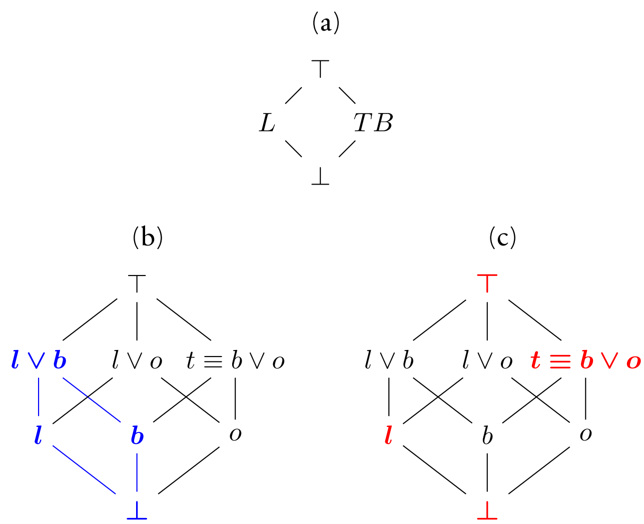
I don't mean to hide behind my new formalism here. The whole theory of probability is built on Boolean algebras, and these just aren't the kind of structures that can represent the kind of distinct-but-equivalent propositions that Mahtani gestures at. One could enhance the structure in some way, as others have done when considering the (hyper)intensionality of belief (Chalmers, 2011). But I don't think that is required here. As I will show, we can capture the sense in which the original TENANT proposition "splits" when your awareness grows in my framework. I leave it to proponents of a more complex model to show what mine is missing.

Propositional identity across awareness change is here modelled by the lattice embedding. 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. 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, where that double label reflects what I said above. Figure 2(b) and (c) show the new algebra Ξ, in which there's a proposition TENANT (t) with propositions BOB (b) and OTHER (o) entailing it. 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 via an embedding.

Here is one possible embedding, which I will label $h : \Omega \to \Xi$. h maps L to l, and TB to 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: $h(\bot) = h(L \wedge TB) = h(L) \wedge h(TB) = l \wedge b = \bot'$. 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 became aware of the possibility of a second tenant. 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. The image of Ω in Ξ, under embedding h, is shown in Figure 2(b) in blue. This embedding makes *The Other Tenant* an example of expansion.

Another embedding, which I will label g, 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.

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



We can now reintroduce Reverse Bayesianism. First, note that the definition of rigid extension needs to be updated, so that RB can be applied. Bradley's formulation assumed that a new algebra is formed by taking the (closure of the) union of the old set of propositions $\mathcal{X}$ with the new propositions $\mathcal{E}$. No questions about the identity of the "old" propositions in the new algebra arose, because it was assumed to be obvious that they were the members of $\mathcal{X}$. Now, we will instead make the mapping explicit. Once again, Ω is the algebra representing the agent's initial state of awareness, and Ξ the algebra representing their final state. Between the two we have a lattice embedding $h : \Omega \to \Xi$. Let P be the agent's prior credence function on Ω, and let P^+ be any probability function on Ξ.

> **Rigid Extension (updated).** For any $A, B \in \Omega$, where $P(A) > 0$ and $P(B) > 0$, P^+ is a rigid extension of P to Ξ iff:
>
> $$\frac{P(A)}{P(B)} = \frac{P^+(h(A))}{P^+(h(B))}$$

With this in place, we have transformed *The Other Tenant* from a case in which RB generates the incorrect answer to one in which there are two, quite sensible, Reverse Bayesian prescriptions corresponding to the two two embeddings I introduced above. On h (blue) we get:

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

Whereas on g (red) we we get:

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

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. In this model, 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. Our revised RB principle won't problematically assign the new possibility zero credence, in the way that Mahtani highlighted.[17]

3.4 Choosing an embedding

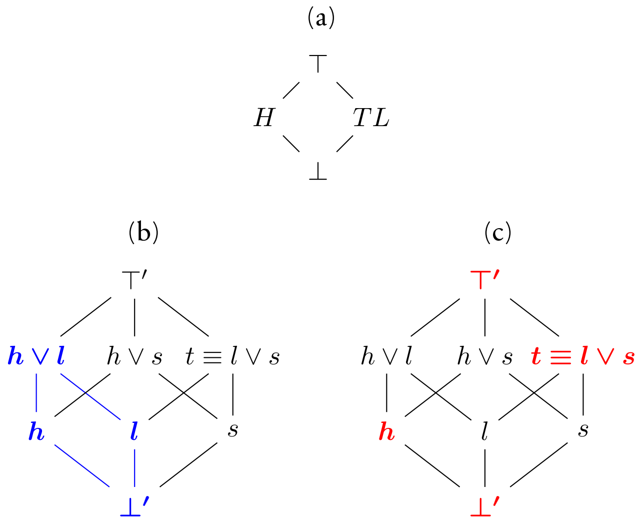
We can now evaluate the two possible embeddings and the prescriptions they generate via Reverse Bayesianism. We would like a systematic way of selecting the right embedding, but to start I will keep it intuitive, informal, and grounded in *The Other Tenant*. The first embedding, h, mapped the TENANT/BOB proposition to BOB, and so RB 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 probabilities for LANDLORD and BOB must be identical: $P^+(l) = P^+(b) = 0.5 - k, P^+(o) = 2k$.

The g embedding, on the other hand, maps TENANT/BOB to TENANT. RB thus requires that the extended probabilities for LANDLORD and TENANT be identical. TENANT is entailed by BOB and OTHER, which are refinements of it. 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: $P^+(l) = P^+(t) = 0.5$. There are no constraints on the credence assigned to b and o so long as $P^+(b) + P^+(o) = 0.5$.

This former assignment seems, to me, to be the more sensible of the two. Mahtani agrees: "given that there might be two tenants, it is natural to suppose that your credence in TENANT should increase relative to LANDLORD" (2020, 9). Taking some credence solely from Bob to give to the other tenant seems unmotivated.

[17]Note that I'm not making any claims about the nature of propositions in this essay. I do not claim that TENANT and BOB *are the same proposition* in a metaphysical sense. I am simply providing a model of an epistemic phenomenon, in which propositions are represented quite crudely, but which I claim is adequate for the purpose of understanding awareness growth and providing rationality considerations for it.

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.



It is gratifying that we can work our way to the right answer in an intuitive way within my framework. But *why* is this the right embedding? I will now introduce a second case, and look for a common pattern between the two. This case is also Mahtani's and has the same structure as *The Other Tenant* but a different intuition about which ratio should be preserved.

> *The Other Tails.* An agent considers a UK 10 pence coin and wonders whether it will land HEADS or TAILS when tossed. They then think about the image on the tails side of the coin. Initially they think all 10p coins have an image of a lion, so that TAILS and LION are equivalent propositions to them. Later, the agent becomes aware that some 10p coins have an image of Stonehenge. So, in their new awareness state, LION and STONEHENGE each entail TAILS. If the agent takes the coin to be fair, how should their initial credences change in light of this new awareness? (paraphrased, from Mahtani, 2020, 9–10)

Once again, we start with two labels, TAILS and LION, for what we will model as one proposition TAILS/LION. When the agent's awareness grows, they end up with a proposition TAILS that has two sub-possibilities, LION and STONEHENGE. There are two possibilities for how to embed the old algebra into the new one. The first embedding, h, maps TAILS/LION to LION and is shown in blue in Figure 3(b). The second, g, maps TAILS/LION to TAILS and is shown in red in Figure 3(c). Once more we can decide which embedding is correct by examining the prescriptions they generate via Reverse Bayesianism.In this case we have a strong intuition that the agent should keep their relative credence in HEADS and TAILS unchanged. All that has happened is that they have become aware of a new image on the tails side of the coin, which we take to be irrelevant to these credences. So in this case, our intuition favours the g embedding. But note that HEADS-TAILS is structurally like LANDLORD-TENANT, and in *The Other Tenant* we concluded that we wanted the LANDLORD-BOB ratio to be preserved. What is the difference?

The Other Tails has a feature that *The Other Tenant* does not: the agent's subjective probabilities for the coin toss are plausibly derived from the chances we commonly take to hold for coin tosses. These chances derive from the symmetry of the coin itself: its two sides and constant weight distribution. These credences 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 we choose an embedding—which is a way of specifying how the agent's awareness has grown—based on their beliefs. In particular we look at the *reasons that their beliefs are structured as they are*, and choose an embedding which allows us to preserve these reasons in the new awareness state.

Can we apply this to *The Other Tenant*? Recall that Mahtani tells us that you start 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. We aren't told any such reasons in the vignette, but we can conjecture: often such judgements of equiprobability are due to reasoning according to a "principle of indifference". Under such reasoning, an agent identifies events which they have no reason to differentiate between, and assigns them equal credence. The principle is notoriously partition dependent, so there will be a way of specifying events *relative to which* they are judged equivalently plausible.

This allows us to distinguish between the {LANDLORD, BOB} partition and the {LANDLORD, TENANT} partition. On the {LANDLORD, BOB} way of thinking, the indifference 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, no way to distinguish their voices, and therefore to assign them equal probability. On the {LANDLORD, TENANT} way of thinking, the reasoning would have to involve these 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 is my proposal: choose the embedding which best preserves the reasoning underlying the agent's initial credal assignments. There is much more to be said in developing this proposal, but I leave that for another time and instead turn my attention to the question of how to revise one's beliefs following awareness growth.

4 Belief Revision

In this section I offer a proposal for belief revision in cases where the agent learns new information as part of their awareness growth experience. I am inspired here by a set of cases that Steele and Stefánsson (2020, n.d.) have recently introduced as challenges for Reverse Bayesianism. The solution to the problems they raise lies partly in the embedding framework developed above, and partly in a two-stage analysis of awareness growth and then belief revision.

4.1 Steele and Stefánsson's evidential relevance cases

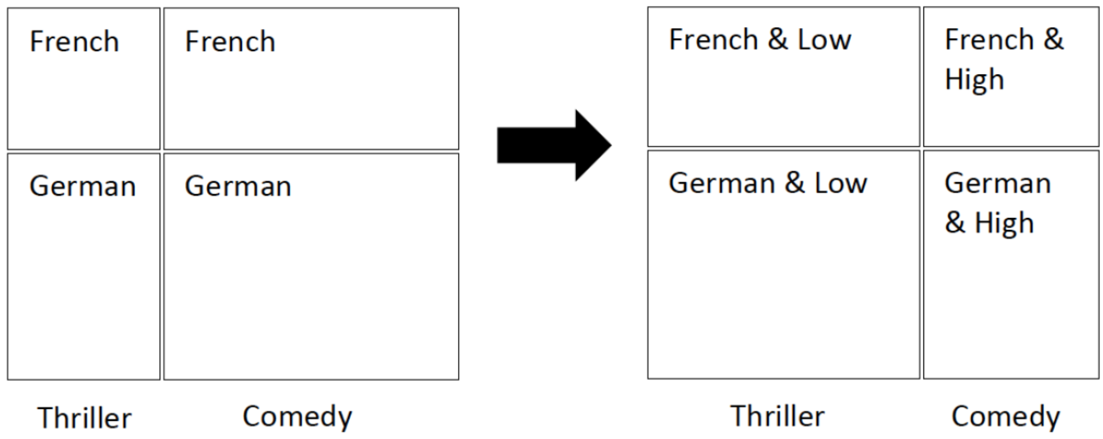
Steele and Stefánsson claim that RB is violated by "examples where awareness grows since an agent becomes aware of a proposition that she takes to be evidentially relevant, intuitively speaking, to the comparison of propositions of which she was already aware" (Steele and Stefánsson, 2020, 11). I will start with a simple case in which all the action is in the evidential relevance, and then move on to a case which also requires the application of my embedding framework.

> *Movies.* Suppose you are deciding whether to see a movie at your local cinema. You know that the movie's predominant 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 it occurs to you that the owner of the cinema is quite 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, 2020, 13–14)

The key to this case is that the agent in *Movies* takes the new property (language difficulty) to be relevant to the probability that the film on show is a thriller. The Reverse Bayesian prescription is that 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 4.

Note that *Movies* 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

Figure 4: The awareness growth in Movies, as Steele and Stefánsson see it (from Steele and Stefánsson, n.d., 65).



it *as if* it were awareness growth, which I agree is possible.[18] But to do so we must treat the "new" information as if it were really new.

Here the "learned" information has two parts. Firstly, "you associate low-level language with thrillers". In the version of this example in their book, Steele and Stefánsson model this as $Q(\text{Thriller} \mid \text{Low})=1$ and $Q(\text{Comedy} \mid \text{High})=1$ (Steele and Stefánsson, n.d., 65). Secondly, since "the owner of the cinema is quite simple-minded, you are, after this realisation, much more confident that the movie will have low-level language than high-level language": $Q(\text{Low}) > Q(\text{High})$. I have represented them as constraints on the agent's posterior credence, foreshadowing how I will model this belief revision step below, but really each is a piece of information that "comes to you" in your remembering.

As with Mahtani's cases, I do not deny that Steele and Stefánsson present the intuitively correct result, here visualised in Figure 4. My proposal is simply that Reverse Bayesianism be regarded as a *partial* specification of the change to the agent's belief state in such cases. Specifically, it deals with the change due *only* to the change of awareness. At it's core is a rule for extending belief from one awareness context to another; it says nothing about how to revise those beliefs in light of new evidence. So it's signature ratio

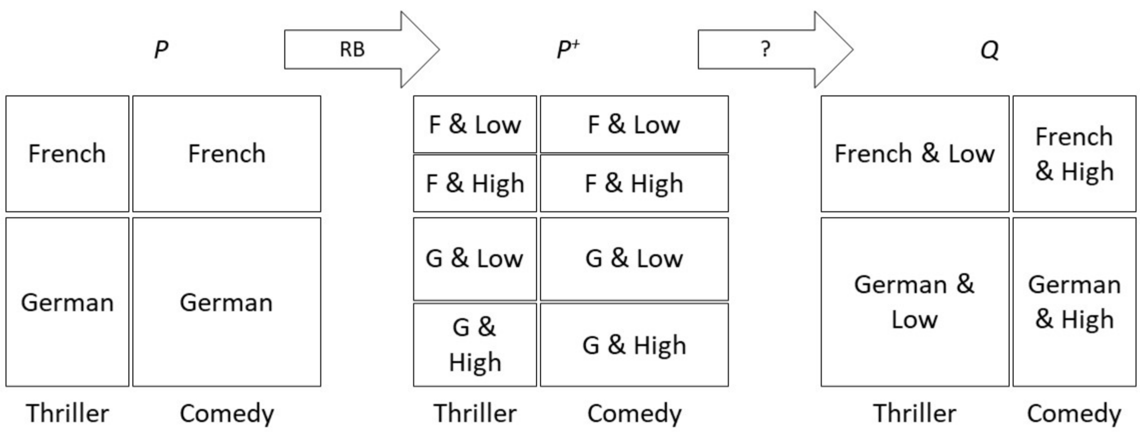
[18] Steele and Stefánsson foresee that a defender of Reverse Bayesianism might argue that this is not a case of genuine awareness growth, and argue that this would be to dodge their challenge. By saying that it can be modelled as if it were true awareness growth, I aim to split the difference with them. I will deal with the case, though I don't accept it as a core instance of the phenomenon of interest.

formula 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 RB applies. Rigidly extending the agent's prior beliefs to the new algebra will preserve the ratio of probabilities of Thriller to Comedy. But that isn't the end of the story. The agent also comes to have Q(Thriller | Low)=1, Q(Comedy| High)=1 and Q(Low) > Q(High). Some belief revision process is required here! But, however this is accomplished, these combine to imply Q(Thriller) > Q(Comedy) as expected. This is summarised in Figure 5.

My proposal has two stages. First, awareness change. The agent apprehends the new propositions and recognises their logical relations to known propositions. In the analysis: we identify the initial and final algebras and then find an embedding to map the old algebra into the new. The agent sees what their prior attitudes "have to say" about the new awareness state. In the analysis: we rigidly extend the agent's probabilities to the new algebra. Second, the agent incorporates any new information they have learned, by updating their new set of beliefs. In the analysis: we update the extended probability to account for the new information, in a procedure I describe in the next section.

Figure 5: Awareness growth followed by belief revision in *Movies*.



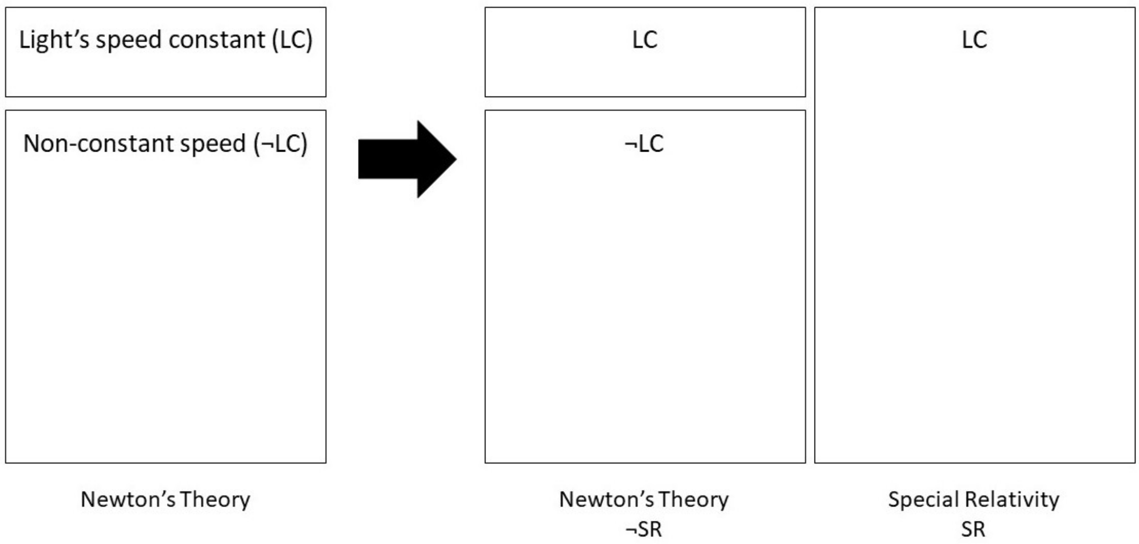
The separation into stages is conceptual rather than temporal: the aim is to focus first on the purely awareness-related aspects of the experience and then to turn to the attitude of belief. The experience that brings about these three steps might (and often will) be a single, unitary experience. The 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: one can't update P with information about the new propositions, as it isn't defined on the right algebra.

In the next section I will sketch a model for generalised belief revision for growing awareness. But first, let us turn to a more complex example, which requires both the embedding framework from above and this two-stage analysis, to draw the pieces of my account together.

> *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 the [SR] changed their relative confidence in such propositions.

As Steele and Stefánsson interpret the example, the agent understands that SR is evidentially relevant to the proposition that 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 $P^+(LC)/P^+(\neg LC) > P(LC)/P(\neg LC)$, 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.

Figure 6: Recreation of Steele and Stefánsson's diagram showing the awareness growth in Relativity, as they see it.



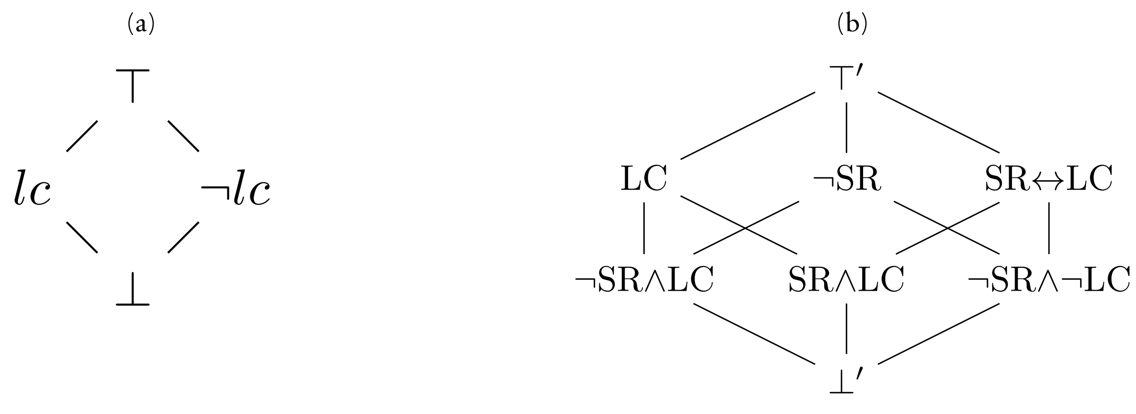
This example is in fact quite different from the above. The constancy of the speed of light is a postulate of Special Relativity, and so SR *logically entails* LC. This is relevant to the lattice structure of the new algebra. As SR entails LC there will not be a SR∧¬LC proposition in the algebra. In the standard model of probabilistic beliefs, we process logical entailment in a different way from evidential relations. The former are modelled in the structure of the underlying algebra, while the latter are modelled in the agent's conditional probabilities.

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 RB cover all the changes that need to take place following awareness growth.

Steele and Stefánsson argue that Bradley's rigid extension does not always result in the Reverse Bayesian ratio formula. and Reverse Bayesianism, write: "[Figure 6] makes vivid that [rigid extension] does not generally entail [the ratio formula], at least for cases of awareness growth by expansion. [Rigid extension] 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" (Steele and Stefánsson, n.d., 64).

On my embedding view, Steele and Stefánsson's application of rigid extension to *Relativity* is incoherent. In order to identify old propositions with new, we need to specify an embedding 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 7: 7(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 ($\top$). In Figure 7(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). As LC is a postulate of SR, I have SR below LC (entailing it) and have labelled SR as SR∧LC to emphasise that there is no SR∧¬LC possibility. On the other hand, there are two ¬SR possibilities: ¬SR∧LC and ¬SR∧¬LC.

Figure 7: Hasse diagrams for the initial algebra for Relativity, and the refined algebra after awareness growth.



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 identify the top element of the initial algebra ($\top$) with "Newton's theory" (¬SR) in the refined algebra *and* identify "light is constant" (lc) with the LC proposition in the new algebra. No lattice embedding can accomplish this. Any embedding h on which $lc \mapsto$ LC will map $\top \mapsto \top'$, because it needs to preserve the ordering of the lattice: $lc \models \top \Rightarrow h(lc) \models h(\top)$. So if I am right that a lattice embedding is the right criterion of identity for awareness changes, Steele and Stefánssonare simply mistaken.

It is possible to map $\top \mapsto \neg$SR, but then "light is constant" must be mapped to the proposition beneath ¬SR: $lc \mapsto \neg$SR∧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: $\top \mapsto \neg$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.[19]

Note again: this has nothing to do with Reverse Bayesianism! It is simply required if you want to have any theory which identifies propositions from the old awareness state with propositions in the new state.

Now on either embedding we can apply Rigid Extension. Due to the structure of the algebra, any positive probability assigned to SR (which is equivalent to SR∧LC in the diagram) will raise the probability of LC. But Rigid Extension simply doesn't have anything to say about what probability one assigns to new propositions such as SR. So there is no conflict. Due to this, *Relativity* doesn't illustrate the point quite in the way Steele and Stefánsson want, because LC is a postulate of Special Relativity, and so the evidential relevance is handled entirely by the structure of the new algebra. 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 7(b) and then to map the old algebra into it with an embedding. These happen "prior" to the action of Reverse Bayesianism, in my multi-stage model. As SR entails LC, any suitable algebra will have the SR proposition

[19]It would be more natural to consider an agent who did not take Newton's theory to be tautologous to begin with. Instead it would be a contingent proposition N, and the agent would entertain the four possible combinations between {N, ¬N} and $\{lc, \neg lc\}$. SR would then be a refinement of ¬N, which would make the analysis more complex. My primary aim here is to show how my modelling approach clears up problems raised by others, in particular Steele and Stefánsson, so I won't develop it.

sitting below it.

This logical relationship does mean that assigning a high probability to SR will increase the probability of LC. But Reverse Bayesianism says nothing to prevent this. It merely says that the starting point for the agent's reasoning should be a rigid extension of their prior beliefs, which (on a suitable embedding) will have the effect of preserving the ratio of credence in LC and ¬LC.

This separation into two stages seems simple enough. What have Steele and Stefánsson to say about it? They reject this separation, on grounds of ad hocness:

> It might be argued that our examples…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 changes to the ratios of probabilities for the old basic propositions. The reason we reject this way of conceiving of the learning events described by our examples is that the two-part structure seems 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. (Steele and Stefánsson, 2020, 15)

The first bit of motivation for this separation is clarity: the two processes are conceptually distinct and come apart in many cases. There are cases of awareness growth (like *Weather* or *The Other Tenant*) that don't involve learning anything about the relevance of the new propositions to the old. Others, like *Movies*, involve learning contingent evidential relations. *Relativity* was different again, with the deductive entailment between a new proposition and an old guaranteeing a certain confirmational structure. Insisting that these all be treated identically in a one-step process seems, to me, unmotivated.

"Spontaneity" seems like the wrong protest about the learning event. There is nothing spontaneous about it—it is brought about by whatever experience leads to the awareness change. The use of examples where what is happening is closer to remembering than true awareness growth may be influencing the guiding intuitions. So to push against that, note that it is quite plausible that one *does* become aware of something before learning about its evidential relations to other matters in many cases—think of awareness growth cases which involve new concepts. Once Naledi becomes aware of snow, when she has grasped the concept, *then* she can appreciate the relation between snow and cold.

I take it that by "hard to rationalise" Steele and Stefánsson mean "hard to provide norms of rationality for". This will depend on the plausibility and normativity of the belief revision procedure in the next section. I suspect that the worry is that this belief revision cannot be accomplished by *Bayesian conditioning*. (Stefánsson has said in presentations that the two-stage approach "saves Reverse Bayesianism at the cost of Bayesianism".) But this is an unreasonable demand! By its very nature, awareness growth cannot be accommodated within a standard Bayesian model. The agents have no priors for the new propositions, and so nothing like a Bayesian update can take place. This much was clear from the outset. So I do not regard it as a failing of the approach that I outline in the next section that it is not Bayesian conditioning. Once more my focus on *true* unawareness is important. In examples like *Movies*, Steele and Stefánsson draw on background knowledge of evidential relations between new propositions and old. As the agent comes to consider a possibility (which they were previously aware of but neglected), they employ this background knowledge. This knowledge is "already there" in a way that puts Bayesianism back into play. But if we want to model such cases as if they were awareness changes, we need to treat the evidential relations as something learned. As this learning event isn't of the Bayesian kind (i.e., it doesn't seem like the change of probability is due to learning a proposition), we need more general tools to handle the belief revision.

Steele and Stefánsson conclude their book length treatment of awareness growth negatively:

> Does rationality (or agential stability) impose any general constraints on the relationship between one's credences prior to and post some growth in awareness? Against the popular position in the philosophy and economics literature, we argued that there are no such general constraints. (Steele and Stefánsson, n.d., 108)

I claim that the two-stage model of Reverse Bayesianism plus generalised belief revision provides those constraints. If the brief passage above about ad hocness is the only reason to reject this in favour of no research programme in this area at all, I expect many will find Steele and Stefánsson overly negative.

4.2 A general model of belief revision

In this section I provide a model of generalised learning that answers the question: How should should agents update their beliefs when they learn new information about the propositions they are newly aware of?

As I noted above, we must start by accepting that learning after (or, as part of) awareness growth cannot 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 is not in Ω and no conditional probabilities $P(\cdot|E)$ are defined. So Bayesians who venture into these waters must give up their familiar comforts.

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. Ideally it should be able to handle quite general inputs: in *Movies* the agent learned an inequality between propositions, and conditional probabilities linking old and new propositions. In order to model this, I will draw on Dietrich et al.'s (2016) generalised belief revision model. They argue that Bayesian learning can be subsumed under an axiomatic framework which also generates other popular updating rules such as Jeffrey and Adams updating. These rules all have the same rational foundations.

Importantly for us, these rational foundations can be applied in awareness growth cases. Roussos (2020) has recently applied this framework to the case of expert deference involving unawareness. Roussos's account of deference allows an agent to adopt an expert's credences about propositions for which the agent has no prior. (Expert testimony is a useful case for thinking about awareness growth, as the new probabilities are precisely stipulated. This allows for clean examples of true unawareness where there is nevertheless a precise constraint on the agent's posterior credences.)

Roussos 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 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, 165). Roussos models all deference cases as providing constraints on posteriors; rather than the agent learning the proposition *that the expert reported their credence as* $P_E(X) = x$, the experience of hearing the report provides the Jeffrey-like constraint $Q(X) = x$ for the agent. These constraints are then used to generate a complete posterior using the Dietrich et al. (2016) generalised belief revision procedure.

For Dietrich et al. (2016), a belief revision rule is a function mapping an initial belief state and an experience to a final belief state. Belief states are sets of probability functions and, as I am interested in belief revision following an extension to a new algebra, we will typically begin with a highly imprecise state P^+. Experiences, or "inputs", are specified extensionally as the set of belief states consistent with the experience. In Jeffrey's example, looking at a piece of cloth in poor light leads an agent to have credence 0.7 that it is green. This input will be modelled as the set of probability functions with $Q(\text{Green}) = 0.7$.[20]

[20]These links back to Richard Jeffrey should not surprise us. Jeffrey was motivated in part by a sense that a realistic agent simply would not have priors for the richly specified experiences he was interested in—see Jeffrey (1983), 160 ff. Dietrich et al. cite Jeffrey's 1957 PhD thesis as the origin of his interest in the idealisation of agential opinionation.

Dietrich et al. (2016) characterise the rationality of belief revision with two conditions: Responsiveness and Conservatism. Responsiveness ensures that the final belief state matches the input (i.e., is drawn from the input set). Conservatism ensures that the belief revision changes nothing that is not required to meet Responsiveness. For a gloss on these notions, let us consider the classic case: Bayesian learning involves an agent learning a proposition E; the input is specified by $Q(E) = 1$, and all events which have inputs of this form are called "Bayes inputs". Bayes' rule takes an input set $\mathbb{Q}_E$ and a prior P, and produced a posterior Q, $Q(X) = P(X|E)$ for all X. Dietrich et al. (2016) use the axioms as follows: Responsiveness demands that Q is in $\mathbb{Q}_E$. Conservatism demands that the belief revision preserves what the input is "silent on" (to be defined below). The experience provides us with no reason to change what it is silent on. In the absence of such a reason, the thinking goes, we should make no change and instead retain our old belief. This is what the axiom of Conservatism says: if an experience is silent on some probability, then that probability should be preserved by the belief revision. In the Bayesian case, this is codified by the Rigidity rule for conditional probabilities: $Q(\cdot|X) = P(\cdot|X), \forall X \in \Omega$. They then show that Bayes' rule is the only way to satisfy Responsiveness and Conservatism for this input type.

This is how the system works in general. An input type is specified. Responsiveness is relatively straightforward: the posterior must be drawn from the input set. Conservatism is more complex: we need a precise definition of silence, and then need to examine what each kind of input is silent on. If the learning experience provides sufficient constraints, Responsiveness and Conservatism generate a unique belief revision rule which produces a unique posterior.

In addition to the classic Bayesian case, Dietrich et al. (2016) look at Jeffrey inputs, which fix the probabilities $Q(A)$ for each proposition A in a partition $\mathcal{A}$. In a longer pre-publication version of the paper they consider the input types for two more rules: Adams conditioning (introduced in Bradley, 2005) and a new rule they call dual-Jeffrey updating (Dietrich et al., 2014). A dual-Jeffrey experience delivers constraints of the form $Q(\cdot|C)$ for each proposition C in a partition $\mathcal{C}$. An Adams experience delivers constraints of the form $Q(A|B)$, for each A and B drawn from two partitions $\mathcal{A}$ and $\mathcal{B}$. In each case, their method identifies what the experience is silent on, and shows that the associated rule is the only Responsive and Conservative rule for that input type. These cases involve experiences which are silent on both conditional and unconditional probabilities, specified relative to the input.

I will briefly explain what it means for a learning experience to be "silent on" the probability of A given B. The intuitive notion they begin with is that $Q(A|B)$ should be unconstrained, and so able to take any value α in $[0, 1]$. But this turns out to be too weak, as it leaves room for a constraint relationship between the conditional probability of A given B and other conditional probabilities. What we need is for the probability of A given B to be unconstrained even after the other parts of Q have been fixed by the experience.

In order to make this precise, we need to know what "the other parts" of Q are. Dietrich et al. decompose Q in terms of A and B, in order to identify the parts of the function which are orthogonal to $Q(A|B)$. They only consider the case $A \models B$, where A, B are in the support of the input $\mathbb{Q}$. Then we can decompose a generic posterior state as follows:

$$
\begin{aligned}
Q(\cdot) &= Q(\cdot|A)Q(A) + Q(\cdot|\neg A)Q(\neg A) \\
&= Q(\cdot|A)Q(A) + Q(\cdot|B\neg A)Q(B \wedge \neg A) + Q(\cdot|\neg B)Q(\neg B) \\
&= Q(\cdot|A)Q(A|B)Q(B) + Q(\cdot|B\neg A)(Q(B) - Q(A|B)Q(B)) + Q(\cdot|\neg B)(1 - Q(B))
\end{aligned}
$$

Thus, Q depends on five factors: $Q(\cdot|A), Q(A|B), Q(B), Q(\cdot|B \wedge \neg A), Q(\cdot|\neg B)$. An input is *silent on* $Q(A|B)$ when this can be freely set once the other four factors are fixed in line with the input.

This sets a procedure we can use in our awareness growth cases. Whatever is learned during the awareness change specifies an input, $\mathbb{Q}$ defined on the new algebra. For each type of input, we identify which probabilities the input is silent on. Now we check to see whether the extended belief, P^+, constrains that probability. (Remember that extensions are generally very imprecise, because all that Reverse Bayesianism tells us to preserve is the probabilities of the embeddings of old propositions, conditional on the embedding of the old top element.) If P^+ does constrain the silent probability, then we preserve that bit of P^+ during the belief revision, under an extension of the logic of Conservatism.

Finding what an input is silent on is difficult and subtle work, as the proofs in Dietrich et al. (2014, 2016) show. But the crucial point for my argument is that there is a way to rationalise even this very general kind of belief revision. Even better: it is the familiar conservatism justification that unifies Bayesian, Jeffrey, Adams and dual-Jeffrey updating.

There are some easy cases, when the learning concerns the new propositions only in a way that doesn't change old probabilities. Then the belief revision and Reverse Bayesianism can fit together very straightforwardly, as the constraints which define the input concern only the *new propositions*, which Reverse Bayesianism has nothing to say about. These include cases when the agent learns new posterior probabilities $Q(E_i)$ for a set of new propositions $\{E_i\}$, or conditional probabilities $Q(E_j|E_i)$, or inequalities between them, etc. For the agent has no priors involving the $\{E_i\}$ propositions, that's part of what it means to say that they underwent a growth of awareness. In such cases we simply find the intersection of P^+ and the input set $\mathbb{Q}$.

Recall the *Movies* case. Here the input is specified by these probabilities: Q(Thriller | Low)=1, Q(Comedy | High)=1 and Q(Low) > Q(High). The last of these is of the "easy" type, as Reverse Bayesianism will not constrain the probabilities of the new propositions Low and High. The first two (the core of the case) form an Adams-type input: a set of conditional probabilities for one partition $\mathcal{A}$ ={Thriller, Comedy}, given another partition $\mathcal{B}$ ={Low, High}. The Conservatism condition for this comes in two parts: the input is silent on the conditional probabilities over the joint partition, and on the unconditional probabilities for the second, conditioned-upon, partition ((Dietrich et al., 2014, 15), see also (Bradley, 2017, 199)):

$$Q(\cdot|AB) = P(\cdot|AB), \forall A, B \in \mathcal{A}, \mathcal{B}$$

$$Q(A) = P(A), \forall A \in \mathcal{A}$$

The uniquely rational update rule for this is Adams conditioning (Bradley, 2017, 197): for any X in the new algebra Ξ:

$$Q(X) = \sum_k \left[\sum_j P^+(X|B_j A_k) \cdot Q(A_j|B_k) \right] P^+(B_k)$$

For clarity, $P^+(X|B_j A_k)$ refers to $P^+(X|$ Thriller, Low) and the other probabilities conditional on that joint partition. $P^+(B_k)$ refers to P^+(Low) and P^+(High). None of these are constrained in P^+. (We might decide to first apply the Q(Low) > Q(High) update, in which case those two are indirectly constrained.) The update is therefore highly imprecise. We can apply the Adams conditioning rule to every member of P^+ and generate a corresponding Q. The result will be a highly imprecise posterior, in which every function obeys the constraints given in *Movies*. Note that this procedure will preserve aspects of P^+ which aren't affected by the *Movies* input! So, in particular, the ratio of probabilities for French and German will be preserved, just as Steele and Stefánsson have it in their diagram, reproduced in Figure 4.

References

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

Richard Bradley. Radical Probabilism and Bayesian Conditioning. *Philosophy of Science*, 72(2):342–364, April 2005. doi: 10.1086/432427.

Richard Bradley. *Decision Theory with a Human Face*. Cambridge University Press, 2017. ISBN 978-1-107-00321-7.

David J. Chalmers. Frege's puzzle and the objects of credence. *Mind*, 120(479):587–635, 2011. doi: 10.1093/mind/fzr046.

B. A. Davey and H. A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, Cambridge, second edition, 2002.

Franz Dietrich, Christian List, and Richard Bradley. A Unified Characterization of Belief-Revision Rules. *unpublished*, page 37, 2014.

Franz Dietrich, Christian List, and Richard Bradley. Belief revision generalized: A joint characterization of Bayes' and Jeffrey's rules. *Journal of Economic Theory*, 162:352–371, 2016. doi: 10.1016/j.jet.2015.11.006.

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

Ellery Eels. Problems of Old Evidence. *Pacific Philosophical Quarterly*, 66:283–302, 1985.

Clark N. Glymour. *Theory and Evidence*. Princeton University Press, Princeton, 1980.

Richard Jeffrey. *The Logic of Decision*. University of Chicago Press, second edition, 1983.

Edi Karni and Marie-Louise Vierø. "Reverse Bayesianism": A Choice-Based Theory of Growing Awareness. *American Economic Review*, 103(7):2790–2810, December 2013. ISSN 0002-8282. doi: 10.1257/aer.103.7.2790.

Edi Karni and Marie-Louise Vierø. Probabilistic sophistication and reverse Bayesianism. *Journal of Risk and Uncertainty*, 50:189–208, 2015.

Edi Karni and Marie-Louise Vierø. Awareness of unawareness: A theory of decision making in the face of ignorance. *Journal of Economic Theory*, 168:301–328, 2017. doi: 10.1016/j.jet.2016.12.011.

Isaac Levi. Subjunctives, Dispositions and Chances. *Synthese*, 34:423–455, 1977.

Isaac Levi. *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance*. MIT Press, 1980.

Anna Mahtani. Awareness growth and dispositional attitudes. *Synthese*, 2020. doi: 10.1007/s11229-020-02611-5.

Joe Roussos. Expert deference as a belief revision schema. *Synthese*, 2020. doi: 10.1007/s11229-020-02942-3.

B.C. Schipper. Awareness. In H. van Ditmarsch, J.H. Halpern, W. van der Hoek, and B. Kooi, editors, *Handbook of Epistemic Logic*, pages 77–146. College Publications, 2015.

A Shimony. Scientific Inference. In R Colodny, editor, *The Nature and Function of Scientific Theories*, pages 79–172. University of Pittsburgh Press, 1970.

Katie Steele and H. Orri Stefánsson. Belief Revision for Growing Awareness. *Mind*, fzaa056, 2020. doi: 10.1093/mind/fzaa056.

Katie Steele and H. Orri Stefánsson. *Beyond Uncertainty*. Cambridge University Press, forthcoming, n.d.

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

S Wenmackers and Jan-Willem Romeijn. New Theory About Old Evidence. *Synthese*, 193 (4):1225–1250, 2016.

P. M. Williams. Bayesian Conditionalisation and the Principle of Minimum Information. *The British Journal for the Philosophy of Science*, 31(2):131–144, 1980. ISSN 0007-0882.