Belief Revision for Growing Awareness

Katie Steele and H. Orri Stefánsson

March 26, 2019

Abstract

The Bayesian maxim for rational learning could be described as *conservative change* from one probabilistic belief function to another in response to new information. This is precisely articulated for the case when we learn that some proposition that we had previously entertained is indeed true (the so-called rule of *conditionalisation*). But can this conservative-change maxim be extended to revising one's beliefs in response to entertaining propositions or concepts of which one was previously unaware? The economists Karni and Vierø (2013, 2015) make a proposal in this spirit. Philosophers have adopted effectively the same rule: revision in response to growing awareness should not affect the relative probabilities of propositions in one's "old" epistemic state. The rule is compelling, but only under the assumptions that its advocates introduce. It is not a general requirement of rationality, or so we argue. We formulate and defend a more restricted principle that we take to be the proper extension of Bayesianism to situations of growing awareness. Moreover, our general model allows a deeper understanding, not only of the phenomenon of growing awareness, but also of the more standard aspects of the Bayesian model.

1 Introduction

Central to *Bayesianism* is the idea that beliefs come in varying degrees (often dubbed *credences*) and that a rational agent's beliefs over the "possibility space", to put it roughly, can be measured in terms of a probability function. This model of belief, also known as *probabilism*, is mathematically elegant and powerful, all the more so given its celebrated intimate connection to rational preference, at least as defined by the so-called *expected utility* axioms. Moreover, the model promises a simple account of learning: it is a question of how the new probabilistic belief function, in light of the learning experience, should relate to the old probabilistic belief function.

The standard Bayesian answer to this question of learning is *conditionalisation*. The rule dictates precisely how to revise beliefs when one gets a particular type of

information, namely, the information that some proposition that one had previously entertained is indeed true. The rule is conservative: change as little as possible in response to the new constraint on one's beliefs, where this is taken to mean that the agent's new probability function should equate to his/her old probability function, conditional on the proposition that is now found to be true. Indeed, "conservative belief change" may be seen as the underlying Bayesian maxim for rational learning, and it might be pitched as "proportioning one's *change* in belief to the *new* evidence".[1] The maxim has been extended to cases where the information one gains is that some proposition is more or less probable than one had thought (in the form of the rule known as *Jeffrey conditionalisation*).[2]

But there are ways in which one can become "better informed" that fall outside the scope of the traditional Bayesian conception of learning and its more recent extensions. In particular, one can become better informed by entertaining propositions or concepts of which one was previously unaware. This is not a learning experience that can be characterised in the usual way: as a constraint on the agent's probability function over a given possibility space. It is rather a revision of this very possibility space. To be fair, standard conditionalisation is often depicted as a change in the possibility space; it is a shrinking of the space that has positive probability. But here we are talking about a more radical (and yet intuitively common) change, in that there is a growth in the agent's conceptual awareness.

Although learning experiences involving conceptual growth have not been central to the Bayesian project, there is no apparent reason why the maxims of probabilism and conservative belief change cannot be extended to these kinds of cases. Indeed, in just that spirit, the economists Karni and Vierø (2013, 2015) propose an extension of the Bayesian model to capture growing awareness, and they defend the following rule for the associated change in belief: When a person becomes aware of new possibilities, she should update her beliefs "in such a way that likelihood [probability] ratios of events in the original [epistemic] state space remain intact" (2013: 2801). They dub this rule *Reverse Bayesianism*. Philosophers too have explored the challenge that growing awareness poses for the traditional Bayesian model, typically under the guise of "the problem of new theories" (Earman 1992; the problem was originally raised by Glymour 1980 as the counterpart to "the problem of old evidence). The most detailed and well-worked-out of the philosophers' responses to this problem (cited below) endorse what is effectively Reverse Bayesianism.

[1] This is of course a play on Hume's prescription in his *Enquiry Concerning Human Understanding* that "A wise man proportions his belief to the evidence."

[2] The maxim has also been extended to cases where the information one gains affects one's conditional probabilities (in the form of the rule known as *Adams conditionalisation*) (Bradley 2005).

We think this is an important development of the popular Bayesian approach to belief and learning. But the extant proposals all suffer from a common limitation: they countenance only certain kinds of awareness growth. Karni and Vierø, for instance, employ an Anscombe and Aumann (1963) underlying framework, as opposed to the more general Jeffrey (1965)-inspired propositional framework that philosophers tend to favour. The Anscombe and Aumann framework (and Savage's 1954 framework, on which Anscombe and Aumann's is based) consists of acts, maximally specific consequences, and states amounting to act-consequence pairs. In Karni and Vierø's model, awareness growth involves either the recognition of new consequences or new acts, and these newly discovered consequences/acts are by stipulation inconsistent with the consequences/acts of which the agent was previously aware. In contrast, the propositional framework treats acts, consequences and states similarly (they are all just propositions), and so allows for cases where an agent becomes aware of propositions that are consistent with the propositions of which she was already aware.[3] As mentioned, philosophers typically appeal to the more general propositional framework, but the existing treatments of growing awareness introduce limitations on the kinds of awareness growth that may be experienced. Wenmackers and Romeijn (2016), for instance, implicitly assume that the propositions that the agent becomes aware of are inconsistent with those of which she was previously aware, and so does Bradley (2017) in his treatment of awareness growth by *expansion* (more on which below). This at least partly explains why these authors defend what is effectively Karni and Vierø's Reverse Bayesianism.[4]

As far as we know, the question of whether so-called Reverse Bayesianism is a general requirement of rationality—whether it is an appropriate learning rule for all kinds of awareness growth—has not been explored. Here we defend a negative answer and so reveal the importance of this question: We argue that the principle defended by Karni and Vierø is not a general requirement of rationality. We go on to specify what we take to be the proper extension of Bayesianism to situations of growing awareness: a restricted version of Karni and Vierø's principle, which we dub *Restricted Reverse Bayesianism*. Our general treatment of awareness growth has another important upshot, in that it draws attention to, and calls for a reconsideration of, the more standard aspects of the Bayesian model. In particular, our model invites reconsideration of the nature of the agent's possibility space at any given time, whether

[3]See e.g. Joyce (1999) for a discussion of how to translate between Jeffrey's propositional framework and Savage's.

[4]Henderson et al. (2010) also offer a detailed response to the problem of new theories (or growing awareness); their response is different, but it too is limited in that they only countenance awareness growth whereby propositions the agent becomes aware of are instances of, and so entail, one of those of which she was previously aware. They too effectively endorse Reverse Bayesianism.

or not the agent proceeds to experience a growth in awareness.

2 Unawareness and traditional Bayesianism

Let's state the Bayesian framework more precisely as well as the problem that unawareness raises. Let P be the function representing the degrees of belief of our agent, defined on "the set of propositions about which the agent has an opinion ... (which) forms an algebra $\mathcal{F}$" (Pettigrew 2011). The algebra $\mathcal{F}$ has the following character:

- $\mathcal{F}$ contains a contradictory proposition ($\bot$).
- $\mathcal{F}$ contains a tautologous proposition ($\top$).
- $\mathcal{F}$ is closed under disjunction, conjunction, and negation. That is, if A and B are in $\mathcal{F}$, then $A \vee B$, $A\&B$ and $\neg A$ and $\neg B$ are also in $\mathcal{F}$.

The Bayesian norms of rationality are taken to establish that P is a probability function.[5] This core Bayesian commitment is known as *probabilism*.[6]

Pettigrew articulates the common assumption that the algebra, $\mathcal{F}$, is best interpreted in terms of (objective) *possible worlds*, understood to be "maximally specific ways the world might be". Specifically, propositions are identified (or at least associated) with the set of those possible worlds that make the proposition true.[7] The tautologous proposition is thus identified with the set of all possible worlds, while the contradictory proposition is identified with the empty set of possible worlds. This is an appealing way to conceive of propositions because it makes vivid how the total probability "mass" is distributed such that the probability for both simple and complex propositions can be derived. Later, Pettigrew allows that for most purposes, it is not useful to quantify over *all* possible worlds; rather, it makes sense to simply refer to *the possible worlds relative to* $\mathcal{F}$, which are specific enough just to assign truth values to each of the propositions in $\mathcal{F}$, and thus as a set amount to a coarsening of the set of possible worlds.

[5] That is, $P(A) \in [0,1]$ for all A in $\mathcal{F}$; $P(\bot) = 0$; $P(\top) = 1$; $P(A \vee B) = P(A) + P(B)$ for all mutually exclusive A and B in $\mathcal{F}$. There are various ways that Bayesians defend the position that a rational agent's belief function is a probability function. Some take the core rationality norms to be pragmatic in character, constraining preferences in the first instance, and only indirectly beliefs, while others take the rationality norms to directly constrain beliefs.

[6] We will not here consider conservative generalisations of probabilism, such as the position that rational belief is representable by a *set* of probability functions, and not necessarily a single precise probability function.

[7] If propositions are identified with sets of possible worlds, the reader might wonder why we characterised the above algebra sententially rather than set-theoretically. The reason is that we want to leave open the possibility that propositions are not sets of possible worlds (see e.g. footnote 16).

Pettigrew does not address the further issue of whether it makes sense to conceive of the agent's possibility space as having the objective status of the set of possible worlds, whether coarsened or not. Indeed, the phenomenon of awareness growth calls for a reconsideration of this standard presumption. It is far from being a benign presumption because, strictly speaking, it means the agent has substantive knowledge of how propositions relate to one another, and that she in some sense grasps the full set of possibilities. But we will get to this issue in due course. For now what is important is simply that the agent, at a given time, has degrees of belief over propositions about which she has opinions, where the set of these propositions forms an algebra $\mathcal{F}$. An account must be given of how the agent understands this algebra. The orthodoxy is to associate the propositions with sets of objective possible worlds, or coarsenings thereof. Later we will offer a different understanding of how an agent understands $\mathcal{F}$ at a time, one that allows us to accommodate the unorthodox kind of belief revision that is our focus in this paper.

Now, Bayesianism tells an agent, whose degrees of belief are represented by P, precisely how she ought to revise these degrees when she learns that some proposition A in $\mathcal{F}$ is true. The norm of conditionalisation—one of the core theses of Bayesian epistemology—requires that for any proposition B, the agent's confidence[8] in B, after learning A (and nothing stronger), should equal her (prior) conditional degree of belief in B given A, i.e., $P(B \mid A)$, which, according to the standard definition of conditional probabilities, equates to $P(A\&B)/P(A)$ whenever $P(A) > 0$. More formally, let P_A represent our agent's degrees of belief after she has learned A. Then the norm of Conditionalisation states that:

Conditionalisation. *For any $A, B \in \mathcal{F}$ and according to any rational agent:*

$$P_A(B) = P(B \mid A)$$

Given the standard definition of conditional probabilities stated above, Conditionalisation is logically equivalent to the conjunction of the following two principles:

Rigidity. $P_A(B \mid A) = P(B \mid A)$

Certainty. $P_A(A) = 1$

Informally, Rigidity says that whatever proposition the agent may learn, her degrees of belief conditional on this proposition should be rigid, or unaffected by the learning experience. Certainty on the other hand says that one is certain of whatever one has learned. The latter condition does not fit well with the intuitive notion of "learning",

[8]We will use "degree of belief" and "confidence" interchangeably throughout this paper.

according to which one could take oneself to have learned something without having become certain of some proposition. Fortunately, the Bayesian framework can be straightforwardly extended to learning experiences where an agent does not learn anything with certainty, without giving up Rigidity, as Richard Jeffrey (1965) proposed. But Jeffrey's extension does not help in solving the problem that is the topic of this paper.

In sum, the Bayesian approach to belief revision can be conceived as two steps that occur in sequence, dubbed "perturbation" and "propagation".[9] The first "perturbation" step is the initial learning experience, while the second "propagation" step is the impact of the learning experience on the the agent's beliefs in other propositions, i.e., on her entire belief function. The Bayesian approach to propagation is moreover one of conservative change, i.e., the perturbation should minimally change the agent's beliefs. For the standard perturbations that Bayesians have considered, conservative propagation seems aptly formalised in terms of the Rigidity condition.

But traditional Bayesianism is silent about perturbations that amount to a growth in awareness. How are such perturbations best characterised? And how should their conservative propagation be formalised as a learning rule? These are the questions we address here. Intuitively, awareness can grow in (at least) two ways. The first of these is intuitively understood as an *expansion* of the possibility space that generates the algebra over which the person's subjective probability function is defined. The second is intuitively understood as a *refinement* of the possibility space that generates the algebra over which the person's subjective probability function is defined. The next section discusses an example that informally illustrates these two intuitive types of awareness growth.

3 Growing awareness

3.1 Informal examples

Suppose you are deciding whether to see a movie at your local cinema. You know that the cinema only shows one "international" movie each evening, but let's suppose that you have no way of checking which movie will be shown tonight without paying the theatre a visit. Moreover, you know that the movie's language and genre will affect your viewing experience. The languages you consider are French and German and the genres you consider are thrillers and comedies. But then you realise that, due to your poor French and German skills, your enjoyment of the movie will also depend

[9] These expressions come from Bradley (2008).

on the level of difficulty of the language, whether high difficulty, or low difficulty. We can understand this as a move from a situation where your state of awareness is represented by Table 1 to one where your state of awareness is represented by Table 2.[10]

	Thriller	Comedy
French	French & Thriller	French & Comedy
German	German & Thriller	German & Comedy

Table 1: Less aware state

	Thriller	Comedy
French & High	French & Thriller & High	French & Comedy & High
French & Low	French & Thriller & Low	French & Comedy & Low
German & High	German & Thriller & High	German & Comedy & High
German & Low	German & Thriller & Low	German & Comedy & Low

Table 2: Awareness gain by refinement

Now suppose you realise that the genre could be drama. To keep things simple, let's assume you realise this when you are in the state of awareness represented by Table 1. The resulting state of awareness is represented by Table 3.

	Thriller	Comedy	Drama
French	French & Thriller	French & Comedy	French & Drama
German	German & Thriller	German & Comedy	German & Drama

Table 3: Awareness gain by expansion

Note that in the shift from the epistemic state represented by Table 1 to the one represented by Table 2, you have *refined* the possibilities you entertain to accommodate a whole new property: "French & Thriller", "German & Thriller", etc., are refined further depending on the difficulty of the language. More precisely, in the case of refinement, a new partition of the possibility space is introduced.In contrast, in the shift from the epistemic state represented by Table 1 to the one represented by Table 3, you have *extended* the possibilities you entertain: given that the possibilities you originally entertained were all instances of the movie being a thriller or a comedy, the possibility of it being a drama cannot be formulated by refining the original possibilities you entertained. So, here no new partition is introduced. Rather, what was previously

[10] "French" and "German" do not represent potential actions, but rather the propositions that the movie being shown is French/German.

thought to be a partition of the possibility space is recognised in fact not to be exhaustive unless extended with new possibilities.

3.2 Modelling challenges

One might be tempted by the view that all instances of growing awareness should be treated as refinements. The thought would be that every partition of the agent's possibilities (at least implicitly) includes a *catch-all* proposition that represents remaining mutually inconsistent possibilities that the agent cannot articulate (Shimony 1970 was an early advocate of an idea of this kind). Table 4 illustrates this idea, where the '?' entries represent the catch-all(s). This is a natural move if one is committed to the view, discussed earlier, that a rational agent's possibility space, only partially articulated as it may be, amounts to the full set of objective possible worlds.

	Thriller	Comedy	???
French	French & Thriller	French & Comedy	French & ???
German	German & Thriller	German & Comedy	German & ???
??	?? & Thriller	?? & Comedy	?? & ???

Table 4: Less aware state with "catch-all" proposition(s)

Not surprisingly then, philosophers have been attracted to this approach. Indeed, the more and less worked-out proposals in the philosophy literature for accommodating growth in awareness (or new theories) appeal to an explicit catch-all, or else similarly abstract propositions that are place-holders for yet-to-be-properly-articulated propositions/theories.[11] For instance, Maher (1995) assumes that the agent's algebra contains variable propositions for *each* of the yet-to-be-formulated theories, and he moreover assumes that the agent assigns a (non-zero) probability to each such proposition. Henderson et al. (2010) propose something similar, although with the added sophistication that the propositions in the agent's algebra form a hierarchy that remains fixed throughout the investigation, i.e., remains unchanged even when the agent becomes aware of new theories that effectively fill in this hierarchy. Wenmackers and Romeijn (2016) appeal to a single catch-all to account for the negation of all explicit theories to date, an idea that Earman (1992) also considers. The economists Grant and Quiggin (2013) incorporate a catch-all in their model too; it is assigned a probability based on the agent's past experience of limited awareness.

The problem with these proposals is either that they constrain what an agent may later come to be aware of (e.g. Maher op. cit., Henderson et al. op. cit.) or else

[11] Bradley (2017) is an exception.

they appeal to propositions that are so abstract from the agent's point of view that it is unclear why it is useful, or whether it is even cogent, to depict the agent as entertaining these propositions (e.g. Wenmackers and Romeijn op. cit., Grant and Quiggin op. cit.). Indeed, on the latter issue: it would seem that, in order for an agent to make sense of a catch-all, she would need to entertain some ultimate or universal set of possibilities relative to which the catch-all can be defined as the complement of those possibilities she can properly articulate. But it is hard to see how the agent could have access to this universal set of possibilities (which might in fact not even be a coherent notion), given that, by assumption, some of these possibilities cannot be articulated. So, it is hard to see how the catch-all could be well-defined for the agent and hence the appropriate object of a probability.[12] Even if it were well-defined, it is hard to see how an agent could assign a probability to a catch-all—the content of which she is assumed not to know—since it would be impossible for her to know whether her evidence speaks in its favour.[13]

A more promising approach to catch-alls involves a subtle shift in modelling perspective, from that of the agent to that of the modeller. The idea is that the catch-all is accessible only to the modeller, not the agent in question, who is modelled as implicitly assigning the catch-all zero probability.[14] In this way, the catch-all can be well defined. Moreover, we might thereby reconcile the possibility space of the agent (at least her implicit possibility space, or rather, her possibility space from the modeller's point of view) with the full space of objective possible worlds. To take a new example: an agent may be modelled as partitioning the set of fundamental physical theories into "Newton's theory", "Einstein's theory", *plus, implicitly*, "other fundamental physical theory", where the last alternative is assigned zero probability and corresponds to all the remaining possible physical theories, from the modeller's all-seeing perspective. We can thus depict the agent as assigning the same probability to "Einstein's theory" and "the negation of Newton's theory", being unaware of any further alternatives; yet this need not entail that these two propositions are equivalent. Furthermore, if the agent were to become aware of a new fundamental theory, say "Quantum theory", this would not change the overall possibility space; it would merely be a refinement, or rather a *replacement*, of the catch-all with two mutually inconsistent propositions, "Quantum theory" and "(revised) other fundamental physical theory".[15]

So there are certain advantages to modelling awareness growth as a (kind of)

[12] We thank A for suggesting this way of putting the problem.

[13] We thank B for suggesting this way of putting the problem.

[14] We thank R for this suggestion.

[15] As noted, this is not actually a refinement, at least as we define refinements in the next subsection, but rather a case of *replacement* (or *retraction/contraction* of the catch-all, followed by expansion with new propositions).

refinement of the agent's (implicit) possibility space, as perceived by the modeller. But this seems a rather convoluted path to take; a contrived reconciliation of the agent's possibility space with the set of objective possible worlds. For some applications, there may be special reason to capture a wiser, third-person perspective on an agent's limited awareness (see, e.g., Fagin and Halpern 1987). But we are not primarily concerned with this sort of application. (Moreover, even a wise modeller may surely turn out to be fallible, failing at any given point to grasp the full set of objective possible worlds.) We are interested in the concerns or limited perspective of a single agent, and how these concerns change with time.

The way forward, we suggest, is to divorce the *agent's possibilities* from objective possible worlds. While it is useful to depict an agent's epistemic outlook in terms of atomic possibilities that are the ultimate bearers of probability, these atomic possibilities need not be the objective possible worlds that many think give meaning to propositions. Indeed, the Bayesian model need not offer an interpretation of propositions. They can simply go uninterpreted; the agent's possibilities being defined as truth functions over these uninterpreted propositions, or more accurately, over the *basic* or *primitive* propositions. In the next section, we spell out this idea more carefully. Broadly speaking, we offer what economists Heifetz et al. (2006, 2008) refer to as the "subjective state-space" approach to modelling changes in awareness. The details of our model differ from those of Heifetz et al. (2006, 2008), but what is common is that an agent's possibility space is constructed from combinations of the basic propositions of which she is aware, which may change over time. It remains to be seen whether the intuitive differences between changes in awareness discussed above via examples—awareness growth due to refinement versus expansion—are indeed borne out by our formal model.

3.3 Formal model

The aim is to enrich the Bayesian model to allow for distinct states of (un)awareness and changes from one such state to another, or, as we shall call it, changes in *awareness contexts*. As discussed above, we seek to model the limited perspective of the agent, rather than the perspective of a modeller who has a broader view of the agent's limited perspective. Moreover, in order to achieve this, we drop the aspiration of providing a model that describes an agent's epistemic state *as well as* offering an interpretation of propositions. As such, we say that an agent's awareness context is defined by a set $\mathbf{X}$ of *basic* propositions of which she is aware (which we assume to be finite). What we call *basic* propositions are primitive propositions (representing simple facts about the world) that do not involve any logical connectives; so, for instance, "French" and

"Thriller" are basic while "¬French" and "French & Thriller" are not. We have said that the basic propositions are not themselves given an interpretation in our model; they are simply the primitive facts that the agent is aware of. (In other words, any deeper interpretation of these propositions, whether in terms of objective possible worlds or some other kind of structure, is not explicitly modelled here. Moreover, we do not here take a stance on whether propositions should be identified with sets; see footnotes 7 and 16.)

Let the *possibilities* that the agent is aware of be truth functions, ω_i, that return "true/false" for each of the basic propositions.Note that below we will occasionally use $\omega_1, \omega_2, ..., \omega_n$ to denote individual possibilities. The *putative* set of possibilities are all the distinct truth functions that take this form, i.e., effectively all the different combinations of truth values for the basic propositions. This is merely the *putative* or *first-pass* set of possibilities, since some will be deemed inconsistent by the agent (to be explained shortly) and thus excluded from the *real* set of possibilities (as recognised by the agent). We may describe the possibilities in terms of conjunctions of the basic propositions for which the ω_i function in question returns "true". So, in the awareness context represented by Table 1, the possibility {ω_i(French) = true, ω_i(German) = false, ω_i(Thriller) = true, ω_i(Comedy) = false} can be described as "French & Thriller". From now on, we will use this latter way of describing possibilities.

For the set of propositions $\mathbf{X}$, let $\mathbf{W_X}$ be the agent's (real) set of possibilities (recall: a subset of the *putative* set of possibilities, containing only the possibilities that the agent regards as consistent). A possibility is *consistent*, by the agent's lights, if all its conjuncts *could* be true, i.e., if the agent does not take the conjuncts to be mutually inconsistent. What an agent takes to be the set of consistent possibilities will depend on what she regards as partitions of the proposition space (corresponding to properties or categories for which one and only one value can be assumed). For instance, for the agent described by Table 1, one partition of the space is {"French", "German"}, these being the candidate values for the language-type property; a necessary condition for being a consistent possibility, then, is that the conjuncts include one and only one of "French", "German".

So an agent's awareness context $\mathbf{X}$ may be just as well defined in terms of her possibility space, $\mathbf{W_X}$. Any given basic proposition X_i can now be associated with a set of possibilities in $\mathbf{W_X}$: the $\omega_i \in \mathbf{W_X}$ for which the proposition X_i is true. For simplicity,

we refer to this set as $\{X_i\}$.[16,17] We can now also generate a Boolean algebra, $\mathcal{F}_{\mathbf{X}}$, in the usual way: $\neg X_i$ is associated with the set $\mathbf{W_X} \setminus \{X_i\}$, $X_i \vee X_j$ is associated with the set $\{X_i\} \cup \{X_j\}$, and $X_i \& X_j$ is associated with the set $\{X_i\} \cap \{X_j\}$. For reasons that will become apparent shortly, the same proposition can be associated with different sets of possibilities in different awareness contexts. So, more formally, we can think of a proposition as a function from the awareness contexts in which the proposition plays a role to the corresponding sets of possibilities.[18]

For simplicity, we will model only *growth* in awareness over time; our model will not countenance the *shrinking* or *contraction* of awareness over time. Inclusion of the latter possibility would complicate the model and its presentation; moreover, there is a tradition in modelling *rational* belief change to consider only incremental learning (gains in information) rather than forgetting (losses of information).[19] That said, contraction of one's concepts, i.e., contraction of the set $\mathbf{X}$, may in some cases not be due to "forgetting" but rather due to considerations that make it an important aspect of rational learning; we leave further exploration of this phenomenon for future work, however.[20]

Now let us address the dynamics of awareness. We say that the agent's *awareness grows* when the awareness context shifts from $\mathbf{X}$ to $\mathbf{X}^+ = \mathbf{X} \cup \mathbf{X_j}$ where $\mathbf{X_j}$ is the set of propositions that are in $\mathbf{X}^+$ but not in $\mathbf{X}$ (so, $\mathbf{X}$ and $\mathbf{X_j}$ are *disjoint*). Note that by the assumptions we made above, when the awareness context shifts from $\mathbf{X}$ to $\mathbf{X}^+$ there is a corresponding shift from $\mathbf{W_X}$ to $\mathbf{W_{X^+}}$ and from $\mathcal{F}_{\mathbf{X}}$ to $\mathcal{F}_{\mathbf{X}^+}$. Strictly speaking, $\mathbf{W_X}$ and $\mathbf{W_{X^+}}$ do not have any possibilities in common; after all, the possibilities in each are truth functions that have a different number of propositions in their domain. If, however, we allow that the possibilities may be described in terms of the proposition that they are each associated with—the conjunction of all basic propositions for which the function in question returns "true"—then $\mathbf{W_X}$ and $\mathbf{W_{X^+}}$ may in certain cases (as we will see shortly) have possibilities in common.

We can finally characterise the difference between awareness growth by expansion and awareness growth by refinement. But the arguments in the remainder of this paper

[16] We are not here suggesting that the basic propositions are identical to, or defined in terms of, the relevant set of possibilities. After all, the possibilities were themselves constructed from propositions that had some prior meaning. One can retain the traditional notion of propositions being identified with sets of objective possible worlds, as per, e.g. Stalnaker (1984), although this is not explicitly represented in our model. The relation of "association" that we appeal to here is intended to be weaker than "identity".

[17] Strictly speaking, the set in question should be thought of as being indexed to the relevant awareness context. If we wanted to make the index explicit, we could, for instance, write $\{X_i\}_{\mathbf{X}}$. But to simplify the notation, we omit making the index explicit.

[18] To clarify: For awareness contexts where the proposition does not play a role, it is not associated with any set of possibilities.

[19] For a notable exception, see Titelbaum (2012).

[20] We thank Z for this suggestion.

hold for all awareness growth, whether it is expansion or refinement. That said, let us measure the *length* of a possibility by the number of propositions for which the function in question returns "true". (Recall that we assume that the set of basic propositions is finite.) We say that the awareness growth was (purely) due to *refinement* if the *number* of possibilities in $\mathbf{W}_{\mathbf{X}^+}$ is greater than in $\mathbf{W}_{\mathbf{X}}$, and moreover, all possibilities in $\mathbf{W}_{\mathbf{X}^+}$ are *longer* (in the sense just described) than in $\mathbf{W}_{\mathbf{X}}$. In contrast, we say that the awareness growth was (purely) due to *expansion* if the *number* of possibilities in $\mathbf{W}_{\mathbf{X}^+}$ is greater than in $\mathbf{W}_{\mathbf{X}}$, without any possibilities becoming *longer* in the sense given. Moreover, in the case of pure refinement, there are no possibilities common to $\mathbf{W}_{\mathbf{X}}$ and $\mathbf{W}_{\mathbf{X}^+}$, while in the case of pure expansion there are some possibilities common to $\mathbf{W}_{\mathbf{X}}$ and $\mathbf{W}_{\mathbf{X}^+}$.[21]

It turns out later, for reasons of simplicity, that we require a single episode of awareness growth to be either pure refinement or else pure expansion. This ensures that the basic propositions in $\mathbf{X}_{\mathbf{j}}$ (those new basic propositions the agent becomes aware of) are mutually inconsistent, at least by the agent's lights. This would not be the case if the growth in question were simultaneously an expansion and a refinement. For instance, in the movie example from above, the basic propositions you become aware of in the expansion case is that the movie shown is a drama, which is consistent with the propositions you become aware of in the refinement case, i.e., that the language level is high/low. Where an agent experiences awareness growth that is a mixture of expansion and refinement, this must be modelled as two sequential episodes of awareness growth.

Return again to our movie example, and suppose that in the least-aware context (Table 1), the only possibilities that the agent is aware of and considers consistent can be characterised as: "French & Thriller", "French & Comedy", "German & Thriller", "German & Comedy". In other words, she regards any possibility that involves "French & German", and likewise "Thriller & Comedy", inconsistent. Now, when awareness grows due to refinement into high and low level language—as represented by the shift from the awareness context represented by Table 1 to the one represented by Table 2—the new possibilities are longer: "French & Thriller & High", "French & Thriller & Low", etc. The number of possibilities also grows, since e.g., "French & Thriller" becomes "French & Thriller & High", "French & Thriller & Low".

In contrast, when awareness grows due to an expansion, e.g., when the agent becomes aware of the possibility of the movie being a drama which she takes to be

[21] This is not the only way to characterise the difference between refinement and expansion. Here is an alternative: in the case of expansion, but not refinement, there is a complex proposition in $\mathcal{F}_{\mathbf{X}}$ that corresponds to the tautology in that it is associated with the full set of possibilities $\mathbf{W}_{\mathbf{X}}$, and yet this proposition does not correspond to the tautology in $\mathcal{F}_{\mathbf{X}^+}$ but is rather associated with a proper subset of $\mathbf{W}_{\mathbf{X}^+}$.

inconsistent with the movie being either a comedy or a thriller—as represented by the shift from the awareness context represented by Table 1 to the one represented by Table 3—the possibilities do not become longer: we simply add "French & Drama", "German & Drama" to the first four possibilities.

4 Reverse Bayesianism

Now that we have sketched our model of awareness change, let us consider the question of how such change should affect the agent's subjective probabilities. Note that we are indeed assuming, in line with the Bayesian maxim of probabilism, that a rational agent's beliefs *for any given awareness context* must satisfy the probability calculus. A brief comment is in order here: The tautology, which has probability one according to the probability axioms (recall footnote 5), must be interpreted such that it depends on the awareness context $\mathbf{X}$: it is associated with the set of all possibilities in that context, $\mathbf{W_X}$, which corresponds, for instance, to $A \vee \neg A$, for any A in $\mathcal{F}_\mathbf{X}$.

We turn now to whether and how an agent's subjective probability function for one awareness context constrains or relates to her subjective probability function once she has experienced a growth in awareness. For instance, suppose again that having found yourself in the epistemic situation represented by Table 1, you become aware that the movie showing could be a drama; perhaps because you come across movie reviews that concern dramas. How should you revise your confidence in the various other propositions in light of this expansion? Or, instead, suppose that in the situation represented by Table 1, you realise that the level of language difficulty further distinguishes possibilities. How should this refinement affect your confidence in other propositions?

Traditional Bayesianism is silent on these two questions. As we have seen, this is not a type of learning experience that the traditional Bayesian framework incorporates. But recently, Karni and Vierø have defended a unified answer to these two questions (at least for the particular kind of decision problem and awareness growth that they represent) in the form of a principle that they call "Reverse Bayesianism".

Let us state Reverse Bayesianism as if it were a general principle transcending the particular type of decision model formulated by Karni and Vierø. We use P (P^+) to represent the probabilistic degrees of belief of the agent before (after) awareness grows. Recall that $\mathbf{X}$ ($\mathbf{X}^+$) are the basic propositions of which the agent is aware before (after) awareness grows. Reverse Bayesianism holds that the ratio between (and, by implication, the ranking of) the probabilities of any two (basic) propositions in the old epistemic state should not change when awareness grows. More formally:

Reverse Bayesianism. *For any $A, B \in \mathbf{X}$ and according to any rational agent:*

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

Versions of this principle have more recently been endorsed by Wenmackers and Romeijn (2016) and Bradley (2017).

Consider what Reverse Bayesianism requires in the movie example we have been discussing. Suppose you find "German" to be twice as likely as "French" before realising that the movie could be a drama. Then after this realisation, you should still find "German" to be twice as likely as "French". Similarly, after you realise that both French and German can be either at a high or a low level of language difficulty, you should still find "German" to be twice as likely as "French". On the face of it, these implications of Reverse Bayesianism seem quite intuitive. For why should the prospect of the movie being a drama, or the fact that the language could have a high or low level of difficulty, change how the prospect of it being a German movie compares, in terms of one's confidence, to the prospect of it being a French movie?

One might suppose that Reverse Bayesianism is intuitively compelling because it precisely captures conservative belief change for the learning experience in question—awareness growth. Indeed, its defenders take it to be the consequence of something akin to the Rigidity condition for this kind of learning perturbation. Bradley, for instance, says as much:

> Within the Bayesian framework, conservation of the agent's relational beliefs is ensured by the rigidity of her conditional probabilities. So we can conclude that conservative belief change [when faced with growing awareness] requires [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. (2017: 229)

Wenmackers and Romeijn similarly suggest that the conservation of "probability ratios among the old hypotheses" follows from the relevant conditional probabilities remaining constant:

> In analogy with Bayes rule, one natural conservativity constraint is that the new [i.e., more aware] probability distribution must respect the old [i.e., less aware] distribution on the preexisting parts of the algebra [i.e., on the distributions' shared domain]. (2016: 1235)

Karni and Vierø also appeal to the constancy of conditional attitudes by way of defending Reverse Bayesianism. In the behaviourist economics tradition, they appeal to

constraints on preferences, and only indirectly on beliefs:

> . . . as the decision-maker's awareness of consequences grows and his state space expands, his preference relation conditional on the prior state space remains unchanged. (2013: 2801)

The above defences of Reverse Bayesianism are arguably sound given the models of awareness growth to which they pertain. As mentioned in the introductory section, however, these models place limitations on the kinds of awareness growth that may be experienced. As a result, these authors' theories of awareness growth are not quite as general as ours. For Karni and Vierø, in cases of awareness growth by expansion, the agent comes to be aware of consequences that are by their very nature inconsistent with the consequences that define her old awareness context.[22] Wenmackers and Romeijn (2016) focus not on act-consequences but rather on scientific theories, but these are similarly assumed to be mutually inconsistent. Bradley's model is more general, but he too, in his endorsement of Reverse Bayesianism, builds in the assumption that, in cases of awareness growth by expansion, the propositions of which an agent comes to be aware are inconsistent with the propositions of which she was previously aware:

> . . . the key to conservative attitude change in cases where we become aware of prospects *that are inconsistent with those that we previously took into consideration* is that we should extend our relational attitudes to the new set in such a way as to conserve all prior relational beliefs . . . (2017: 229, emphasis added)

Moreover, we hold that Bradley implicitly assumes only "vanilla" kinds of awareness growth by refinement (as will be made clearer below).

We allow that Reverse Bayesianism is defensible for the limited kinds of awareness growth that the above authors consider. But the question remains as to whether this learning rule is defensible in a more general and carefully spelled-out model as per section 3.3 above; i.e., a model in which the propositions the agent becomes aware of may be consistent with those she is already aware of, and where the underlying possibility space is not fixed (by the set of objective possible worlds), but rather changes as awareness grows. For starters, it is not so obvious how to specify a Bayesian-style rigidity condition in a model like ours for which, strictly speaking, different awareness contexts have no possibilities in common. And then there is the question of whether the appropriate version of Rigidity entails Reverse Bayesianism. But more importantly,

[22] Karni and Vierø's treatment of awareness growth by refinement involves awareness of new acts, as opposed to new consequences, and is somewhat more complicated; here we focus primarily on their treatment of awareness growth by expansion.

Reverse Bayesianism may not even be the right result for belief change in the more general model, as we will see in the next section.

5 Counterexamples to Reverse Bayesianism

It is not hard to see that Reverse Bayesianism cannot generally be true once we move beyond the constrained models of its defenders. That is, one can devise examples where Reverse Bayesianism is violated without irrationality on behalf of the agent in question. All we need are examples where awareness grows since an agent becomes aware of a proposition that she takes to be evidentially relevant to the comparison of propositions of which she was already aware. For in that case, the ratio between probabilities of propositions of which the agent was already aware will not stay the same; one will become more probable compared to the other, just like in ordinary cases where one learns evidence relevant to the comparison of hypotheses.

In fact, the history of science is full of examples that undermine Reverse Bayesianism, for the above reason. Here is a particularly prominent such example:

Example 1. *Nineteenth century physicists were unaware of the General Theory of Relativity (GTR). That is, not only did they not take the theory to be true; they had not even entertained the theory. However, they had entertained various propositions that they regarded as evidentially relevant to the theory once Einstein brought it 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 will always be measured at 300,000 km/s independently of how fast the investigator is moving. Therefore, the awareness and subsequent acceptance of the GTR did in fact, and for good reason, change their relative confidence in such propositions.*

Not all examples where Reverse Bayesianism fails come from the history of science. Here is a more mundane, or everyday, example:

Example 2. *Suppose you happen to see your partner enter your best friend's house on an evening when your partner had told you she would have to work late. At that point, you become convinced that your partner and best friend are having an affair. You discuss your suspicion with another friend of yours, who points out that perhaps they were meeting to plan a surprise party to celebrate your upcoming birthday—a possibility that you had not even entertained. Becoming aware of this possible explanation for your partner's behaviour makes you doubt that she is having an affair, indeed it lowers your confidence that she is having an affair relative to other propositions that you entertained, such as her job being overly demanding.*

The relevant feature that the above examples share is that there are some basic propositions that the agents of interest had entertained before awareness grew—respectively the proposition that the speed of light will always be measured at 300,000 km/s independently of how fast the investigator is moving and the proposition that the agent's partner is having an affair—that become more or less probable in comparison to other "old" basic propositions when awareness grows, i.e., respectively when the GTR and the possibility of a surprise party is brought to the agent's attention. For instance, the hypothesis that the measure of the speed of light is constant may become much less probable relative to, say, a hypothesis about gravity, once awareness grows to include the GTR. And the partner having an affair may become much less probable relative to, say, the proposition that her job is overly demanding, once awareness grows to include the possibility that she is planning a surprise party. But that means that, as awareness grows, the ratios between the probabilities of basic propositions in the agent's old awareness context change, in violation of Reverse Bayesianism. But there is nothing irrational about this violation of Reverse Bayesianism. Therefore, Reverse Bayesianism is not a general constraint on rational belief revision.

A defender of Reverse Bayesianism might argue that the above two examples do not undermine their thesis, since, for instance, the proposition picked out by the sentence "the speed of light will always be measured at 300,000 km/s independently of how fast the investigator is moving" is different before and after the speaker becomes aware of the General Theory of Relativity. (Similarly, the proposition picked out by the sentence "my partner is having an affair" is different before and after the speaker realises that their partner might be organising a surprise party.) The idea is to characterise the old and new propositions in such a way that the new propositions are *not* evidentially relevant (which can be understood informally for now; we spell it out formally in the next section) to the comparison of old basic propositions. For instance, the physics case might be spelled out as follows: despite appearances, the agent's growth in awareness is not simply an expansion of the "fundamental physical theory" partition to include the GTR; there is also an expansion of the "light hypothesis" partition to include the GTR version of the (speed-of-) light hypothesis. As a result, the addition of the GTR has no bearing on the original (speed-of-) light hypothesis relative to other basic propositions, such as, say, the hypothesis concerning gravity, in conformity with Reverse Bayesianism. It might be added that, if the new propositions of which the agent becomes aware *were* apparently evidentially relevant to the basic propositions in the old awareness context, then we would *not* have a case of *genuine* awareness growth, to which Reverse Bayesianism is limited.[23]

[23]The implication is that we would rather have a case of irrational and/or poorly represented belief

This way of saving Reverse Bayesianism however seriously weakens the common-sense appeal and normative interest of the thesis, and seems rather *ad hoc*, as the examples under consideration are surely as genuine cases of awareness growth as any. Moreover, if the aim is to represent the subjective point of view of some agent, then it is surely more natural to take the individuation of propositions at face value, such that, with respect to our example above, the speed-of-light hypothesis corresponds to the same proposition before and after recognition of the General Theory of Relativity. But that means that new propositions may well have a bearing on the relative probabilities of old basic propositions. Better to modify the Reverse Bayesian principle itself (the strategy we pursue in the next section) than to modify what counts as genuine awareness growth.

So, we can conclude that we should not impose Reverse Bayesianism as a general constraint on how a rational agent can revise her degrees of belief when her awareness grows. The above counterexamples, however, both involve what we called awareness growth by *expansion*. But as previously mentioned, proponents also want to impose Reverse Bayesianism as a constraint on how a rational agent can revise her degrees of belief when her awareness grows due to *refinement* (see e.g. Karni and Vierø 2013: 2803). And one might well hope that despite the above counterexamples, the principle could be retained for belief revision due to refinement.

Unfortunately, counterexamples similar to those discussed above also undermine Reverse Bayesianism understood in this latter way. Consider a third example, which is an elaboration on the awareness growth represented by the shift from Table 1 to Table 2 in section 3, a case of awareness growth by refinement:

Example 3. *Suppose you are deciding whether to see a movie at your local cinema. You know that 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.*

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 ordering of some pair of basic propositions ("Thriller",

change.

"Comedy") in the old awareness context. In particular, finding that the property typically takes some particular value ("Low", in the example in question), changes the ordering of (and hence the ratio between the probabilities of) some pair of basic propositions in the old awareness context.

In sum, the above examples show that Reverse Bayesianism cannot hold in full generality, neither as a constraint on belief revision due to expansion nor as a constraint on belief revision due to refinement. In other words, there are examples where awareness growth, either due to refinement or due to expansion, alters the agent's probability ratios for the basic propositions that she was aware of all along, without her thereby displaying any irrationality.

Before moving on, we note a different potential criticism of our analysis. 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 cases, 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 (Bradley 2005) learning event.[24] 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 reject 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.

6 Restricted Reverse Bayesianism

So, Reverse Bayesianism is not a general requirement of rationality, we contend. In particular, it fails with respect to propositions A and B if the awareness growth favours one of these propositions over the other. We need to retreat to a more restricted rule for belief change in cases of growing awareness. One might suppose that the retreat should be to the kind of rigidity principle defenders of Reverse Bayesianism apparently take as fundamental. The problem is simply that the relevant rigidity condition does not entail Reverse Bayesianism when stated in general terms. Or so the argument

[24] We thank X for this suggestion; it also resonates with the discussion of awareness change in Hill (2010).

might go.

The first hurdle, in terms of exploring this position, is to give a precise statement of the relevant rigidity condition for awareness growth. Informally, the idea is that the probabilities of the old propositions, conditional on, roughly speaking, "how things were before", should be rigid or unaffected by what the agent is newly aware of. But how can we characterise "how things were before" in a model where awareness growth results in an entirely new possibility space, i.e., a space constituted by different possibilities? Our solution is to identify the smallest set of possibilities in the new awareness context that corresponds to what used to be the tautology in the old awareness context. So, for instance, with respect to our movie example, when awareness grows by expansion to incorporate the new genre of "Drama", as per the shift from Table 1 to Table 3, the proposition corresponding to the smallest set of new possibilities and which used to be the tautology in the old awareness context is the disjunction of all the old genres, i.e., "Thriller $\vee$ Comedy". In the case of refinement, any such proposition will simply correspond to the set of all possibilities constituting the new awareness context. This allows us to specify a rigidity condition that one might take to be the appropriate extension of Bayesian belief change to the case of growing awareness:

Awareness Rigidity First Pass. *Let T^* in $\mathcal{F}_{\mathbf{X}}$ be the proposition that, amongst those associated with the full set $\mathbf{W}_{\mathbf{X}}$, is associated with the smallest subset of $\mathbf{W}_{\mathbf{X}^+}$. For any rational agent and for any $A \in \mathcal{F}_{\mathbf{X}}$:*

$$P^+(A \mid T^*) = P(A)$$

This, we think, captures the rigidity condition that defenders of Reverse Bayesianism take as more fundamental than Reverse Bayesianism itself. But there are two problems with taking Awareness Rigidity First Pass as the rule for belief change under growing awareness. The first problem is that, for some cases (of refinement), it says too much—it is not plausibly a rationality requirement. The second problem is that, in yet other cases (of expansion), it says too little—there is more to say about rational belief change under growing awareness than what is stated by Awareness Rigidity First Pass.

On the first problem: Awareness Rigidity First Pass does in fact entail Reverse Bayesianism in cases where awareness grows by refinement (since it effectively requires that the probabilities for all propositions in the old awareness context remain unchanged), and we have already argued that Reverse Bayesianism is not plausible even in cases of refinement. So by modus tollens, Awareness Rigidity First Pass is not

a plausible rule for belief change to the extent that it applies to refinement. In order to fix this problem, the rigidity condition needs to be refined so that it only substantially constrains belief change in the case of awareness growth by expansion, and is trivially satisfied in the case of refinement. This is done by restricting the condition to cases of expansion (cases where there is a proposition corresponding to the old awareness context and a proper subset of the new awareness context).

Awareness Rigidity. *Let T^* be the proposition in $\mathcal{F}_{\mathbf{X}}$ that is associated with the full set $\mathbf{W}_{\mathbf{X}}$, and with a proper subset of $\mathbf{W}_{\mathbf{X}^+}$. If there is such a proposition T^*, then, for any rational agent and for any $A \in \mathcal{F}_{\mathbf{X}}$:*

$$P^+(A \mid T^*) = P(A)$$

We see that, while the condition is stated generally, Awareness Rigidity only constrains belief change in cases of awareness growth by expansion, since only in such cases would there be a proposition T^*.

Awareness Rigidity is a plausible constraint on belief change under growing awareness, we claim. We say more about this below. But now let us confront the second problem posed above: the principle is too limited. There is an additional constraint that can be placed on belief change that captures the special cases where the relationship between probability ratios specified by Reverse Bayesianism ought to hold. These are cases where awareness growth is evidentially irrelevant for the comparison of pairs of propositions in the old awareness context. Indeed, we propose the following restricted version of Karni and Vierø's principle, which is not undermined by our counterexamples:[25]

Restricted Reverse Bayesianism (RRB)**.** *For any rational agent, for any $A, B \in \mathcal{F}_{\mathbf{X}}$, if the agent's awareness growth, from $\mathbf{X}$ to $\mathbf{X}^+$, is evidentially irrelevant for A vs. B, then:*

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

RRB does not say very much unless we specify what it means for awareness growth to be "evidentially irrelevant for A vs. B". We propose the following account of evidential irrelevance:

Definition (Evidential irrelevance)**.** *For any $A, B \in \mathcal{F}_{\mathbf{X}}$, we say that an agent's awareness growth, from awareness context $\mathbf{X}$ to $\mathbf{X}^+$, where $\mathbf{X_j}$ is the set of all basic propositions $X_j \in \mathbf{X}^+$*

[25]Note that, given the evidential irrelevance restriction, RRB can be stated with respect to any pair of propositions in $\mathcal{F}_{\mathbf{X}}$; there is no need for a further restriction to pairs of *basic* propositions.

such that $X_j \notin \mathbf{X}$, *is* evidentially irrelevant for A vs. B *whenever:*

$$\textit{either} \quad P^+(A \mid \bigvee_{X_i \in \mathbf{X_j}} X_i) = 0 = P^+(B \mid \bigvee_{X_i \in \mathbf{X_j}} X_i)$$

$$\textit{or} \quad \frac{P^+(A \mid \bigvee_{X_i \in \mathbf{X_j}} X_i)}{P^+(B \mid \bigvee_{X_i \in \mathbf{X_j}} X_i)} = \frac{P(A)}{P(B)}$$

Note that the above condition, evidential irrelevance, does not hold in the counterexamples to Reverse Bayesianism discussed in section 4, since, for instance, the probability ratio (in terms of probability function P^+) of your partner having an affair to her job being overly demanding (or for that matter to her not having an affair), conditional on your partner and your friend meeting to organise a surprise party, is lower than the "old" probability ratio of your partner having an affair to her job being overly demanding (or otherwise not having an affair). Similarly, in the example involving refinement, the probability ratio (in terms of P^+) of the movie being a thriller to its being a comedy, conditional on the disjunction of low- and high-level language, differs from the "old" probability ratio of the movie being a thriller to its being a comedy. In sum, both in the counterexamples to Reverse Bayesianism involving expansion and in the counterexamples involving refinement, evidential irrelevance is violated.

On the other hand, the original examples of awareness growth introduced in section 3.1 *are* intuitively cases of evidential irrelevance (with respect to all pairs of "old" basic propositions), by the above definition. In the expansion case, you become aware that the movie could be a drama. This is irrelevant, for instance, to your relative confidence in the movie being a thriller vs. a comedy, since your probability (in terms of P^+) of the movie being a thriller (or otherwise a comedy) conditional on it being a drama is zero (satisfying the first clause of the evidential irrelevance condition). Thus your relative confidence in the movie being a thriller vs. a comedy is unaffected by the awareness growth. Likewise, the realisation that the movie could be a drama is irrelevant to your relative confidence in the language being French vs. German, since your probability ratio (in terms of P^+) of the language being French to it being German, both conditional on it being a drama, matches your "old" probability ratio of the language being French to it being German (satisfying the second clause of the evidential irrelevance condition). In the original telling of the refinement story, too, evidential irrelevance was satisfied for all pairs of basic propositions. For instance, the probability ratio (in terms of P^+) of the language being French to it being German, both conditional on the disjunction of the language-levels that you come to realise, matches your "old" probability ratio for the language being French to it being German.

Note that Restricted Reverse Bayesianism presupposes Awareness Rigidity. The

constancy of ratios holds for evidential irrelevance, as per the statement of RRB, just in case Awareness Rigidity also holds. For if Awareness Rigidity did not hold, then even though the discovered propositions would not favour, say, A over B, it could still be the case that the awareness growth would favour A over B by changing how the probability of A compares to that of B, *independently of the newly discovered propositions*. So, if Awareness Rigidity did not hold, then *Restricted* Reverse Bayesianism would not hold.

We believe that Awareness Rigidity, as defined above as a substantial constraint on belief change in cases of expansion, is a rather trivial assumption. We cannot think of a good counterexample to it, nor can we think of a plausible mechanism by which it would be violated. Whether or not the awareness growth is such that what one becomes aware of affects how propositions compare in terms of their probability, it is hard to see why the growth would *otherwise* affect how propositions compare in terms of their probability. Such an effect would seem mysterious and hard to rationalise. But if Awareness Rigidity always holds, as we claim, then Reverse Bayesianism holds w.r.t. A and B whenever the growth in awareness is evidentially irrelevant for A vs. B; in other words, then *Restricted* Reverse Bayesianism always holds.

7 Concluding remarks

In conclusion, we believe that the principle we have suggested, Restricted Reverse Bayesianism, is an appropriate extension of Bayesian reasoning. In addition to being untouched by our counterexamples to unrestricted Reverse Bayesianism, the restricted principle has great intuitive plausibility. After all, it is supported by the more general view that one should not change one's relative confidence in A vs. B unless one gains evidence that one takes to be evidentially relevant for A vs. B. Bayesians are already committed to the more general view, which is an instance of the type of "epistemic conservativism" that is embedded in the norm of Conditionalisation. Hence, it is, we contend, appropriate to take Restricted Reverse Bayesianism to be a minimal extension of traditional Bayesianism to situations of growing awareness.

Finally, it is worth noting that our RRB principle does not settle all the details of an agent's "new" probability function after a growth in awareness. But that is as it should be. As Karni and Vierø (2013: 2805) correctly point out, we should not expect Bayesianism—even after it has been extended to cases of growing awareness—to determine what probability, for instance, to assign basic propositions one becomes newly aware of. For that is just an instance of the more general problem of which priors to adopt. And just as the Bayesian theory, as it is typically understood, does not

determine which priors to adopt for the propositions in one's initial state of awareness, nor does it determine which probability to assign a basic proposition once one becomes aware of it.

References

Anscombe, F. J. and R. J. Aumann (1963). A definition of subjective probability. *The Annals of Mathematical Statistics 34*(1), 199–205.

Bradley, R. (2005). Radical probabilism and Bayesian conditioning. *Philosophy of Science 72*(2), 342–364.

Bradley, R. (2008). Preference kinematics. In T. Grüne-Yanoff and S. O. Hansson (Eds.), *Preference Change: Approaches from Philosophy, Economics and Psychology*. Springer.

Bradley, R. (2017). *Decision Theory with a Human Face*. Cambridge University Press.

Earman, J. (1992). *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. MIT Press.

Fagin, R. and J. Y. Halpern (1987). Belief, awareness, and limited reasoning. *Artificial Intelligence 34*(1), 39 – 76.

Glymour, C. (1980). Why I am not a Bayesian. In C. Glymour (Ed.), *Theory and Evidence*. Princeton University Press.

Grant, S. and J. Quiggin (2013). Bounded awareness, heuristics and the precautionary principle. *Journal of Economic Behavior & Organization 93*(C), 17–31.

Heifetz, A., M. Meier, and B. Schipper (2006). Interactive unawareness. *Journal of Economic Theory 130*(1), 78–94.

Heifetz, A., M. Meier, and B. C. Schipper (2008). A canonical model for interactive unawareness. *Games and Economic Behavior 62*(1), 304 – 324.

Henderson, L., N. D. Goodman, J. B. Tenenbaum, and J. F. Woodward (2010). The structure and dynamics of scientific theories: A hierarchical bayesian perspective. *Philosophy of Science 77*(2), 172–200.

Hill, B. (2010). Awareness dynamics. *Journal of Philosophical Logic 39*(2), 113–137.

Jeffrey, R. (1965). *The Logic of Decision*. The University of Chicago Press.

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

Karni, E. and M.-L. Vierø (2013). "Reverse Bayesianism": A choice-based theory of growing awareness. *American Economic Review 103*(7), 2790–2810.

Karni, E. and M.-L. Vierø (2015). Probabilistic sophistication and reverse Bayesianism. *Journal of Risk and Uncertainty 50*(3), 189–208.

Maher, P. (1995). Probabilities for new theories. *Philosophical Studies 77*(1), 103–115.

Pettigrew, R. (2011). Epistemic utility arguments for probabilism. In E. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*.

Savage, L. (1954). *The Foundations of Statistics*. John Wiley & Sons.

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

Stalnaker, R. (1984). *Inquiry*. MIT Press.

Titelbaum, M. G. (2012). *Quitting Certainties: A Bayesian Framework Modeling Degrees of Belief*. Oxford University Press.

Wenmackers, S. and J. Romeijn (2016). New theory about old evidence. *Synthese 193*(4), 1225–1250.