# Awareness Growth in Bayesian Networks

July 28, 2022

*Wordcount (including footnotes)*: 5,392

We examine different counterexamples to Reverse Bayesianism, a popular theory that addresses the problem of awareness growth. We agree with the general skepticism toward Reverse Bayesianism, but submit that the problem of awareness growth cannot be tackled in an algorithmic manner, because subject-matter, structural assumptions need to be made explicit. Thanks to their ability to express probabilistic dependencies, we illustrate how Bayesian networks can help to model awareness growth in the Bayesian framework.

## 1 Introduction

Learning is modeled in the Bayesian framework by the rule of conditionalization. This rule posits that the agent's new degree of belief in a proposition $H$ after a learning experience $E$ should be the same as the agent's old degree of belief in $H$ conditional on $E$. That is,

$$\mathsf{P}^E(H) = \mathsf{P}(H|E),$$

where $\mathsf{P}()$ represents the agent's old degree of belief (before the learning experience $E$) and $\mathsf{P}^E()$ represents the agent's new degree of belief (after the learning experience $E$).

Both $E$ and $H$ belong to the agent's algebra of propositions. This algebra models the agent's awareness state, the propositions taken to be live possibilities. Conditionalization never modifies the algebra and thus makes it impossible for an agent to learn something they have never thought about. Even before learning about $E$, the agent must already have assigned a degree of belief to any proposition conditional on $E$. This picture commits the agent to the specification of their 'total possible future experience' (Howson, 1976), as though learning was confined to an 'initial prison' (Lakatos, 1968).

But, arguably, the learning process is more complex than what conditionalization allows. Not only do we learn that some propositions that we were entertaining are true or false, but we may also learn new propositions that we did not entertain before. Or we may entertain new propositions—without necessarily learning that they are true or false—and this change in awareness may in turn change what we already believe. How should this more complex learning process be modeled by Bayesianism? Call this the problem of awareness growth.

The algebra of propositions need not be so narrowly construed that it only contains propositions that are presently under consideration. The algebra may also contain propositions which, though outside the agent's present consideration, are still the object, perhaps implicitly, of certain dispositions to believe.[1] But even this expanded algebra will have to be revised sooner or later. The algebra of propositions could in principle contain anything that could possibly be conceived, expressed, thought of. Such a rich algebra would not need to change at any point, but this is an implausible model of ordinary agents with bounded resources such as ourselves.

Critics of Bayesianism and sympathizers alike have been discussing the problem of awareness growth under different names for quite some time, at least since the eighties. This problem arises in a number of different contexts, for example, new scientific theories (Chihara, 1987; Earman, 1992; Glymour, 1980), language changes and paradigm shifts (Williamson, 2003), and theories of induction (Zabell, 1992). A proposal that has attracted considerable scholarly attention in recent years is Reverse Bayesianism (Bradley, 2017; Karni & Vierø, 2015; Wenmackers & Romeijn, 2016). The idea is to model awareness growth as a change in the algebra while ensuring that the proportions of probabilities of the propositions shared between the old and new algebra remain the same in a sense to be specified.

Let $\mathscr{F}$ be the initial algebra of propositions and let $\mathscr{F}^+$ the algebra after the agent's awareness state has grown. Both algebras contain the contradictory and tautologous propositions $\bot$ and $\top$, and they are closed under connectives such as disjunction $\vee$, conjunction $\wedge$ and negation $\neg$. Denote by $X$ and $X^+$ the subsets of these algebras that contain only basic propositions, namely those without connectives. **Reverse Bayesianism** posits that the ratio of probabilities for any basic propositions $A$ and $B$ in both $X$ and $X^+$—the basic propositions shared by the old and

[1] Roussos (2021) notes that, for the sake of clarity, the problem of awareness growth should only address propositions which agents are *truly* unaware of (say new scientific theories), not propositions that were temporarily forgotten or set aside. This is a helpful clarification to keep in mind, although the recent literature on the topic does not make a sharp a distinction between true unawareness and temporary unawareness.

new algebra—remain constant through the process of awareness growth:

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

where $\mathsf{P}()$ represents the agent's degree of belief before awareness growth and $\mathsf{P}^{+}()$ represents the agent's degree of belief after awareness growth.

Reverse Bayesianism is an elegant theory that manages to cope with a seemingly intractable problem. As the awareness state of an an agent grows, the agent would prefer not to throw away completely the epistemic work they have done previously. The agent may desire to retain as much of their old degrees of beliefs as possible. Reverse Bayesianism provides a simple recipe to do that. It also coheres with the conservative spirit of Bayesian conditionalization which preserves the old probability distribution conditional on what is learned.

Unfortunately, Reverse Bayesianism does not deliver the intuitive results in all cases. There is no shortage of counterexamples against it in the recent philosophical literature (Mathani, 2020; Steele & Stefánsson, 2021). In addition, attempts to extent traditional arguments in defense of Bayesian conditionalization to the case of awareness growth seem to hold little promise (Pettigrew, forthcoming). If the consensus in the literature is that Reverse Bayesianism is not the right theory of awareness growth, what theory (if any) should replace it?

Here we offer a diagnosis of what is wrong with Reverse Bayesianism and outline an alternative proposal. The problem of awareness growth—we hold—cannot be tackled in an algorithmic manner because subject-matter assumptions, both probabilistic and structural, need to be made explicit. So any theory of awareness growth cannot be a purely formal theory. This does not mean, however, we should give up on probability theory altogether. Thanks to its ability to express probabilistic dependencies, we think that the theory of Bayesian networks can help to model awareness growth in the Bayesian framework. We illustrate this claim as we examine different counterexamples to Reverse Bayesianism.

## 2 Counterexamples to Reverserve Bayesianism

A common set of cases of awareness growth are usually referred to in the literature by the label *awareness expansion*. A precise definition of expansion can be tricky to provide, but a rough characterization will suffice here. Suppose, as is customary, propositions are interpreted as sets of possible worlds, where the set of all possible worlds is the possibility space. Awareness expansion occurs when a new proposition is added to the algebra and its interpretation includes possible worlds not in the original possibility space. So the addition of the new proposition causes the possibility space to expand.

Perhaps the most straightforward example of awareness expansion occurs when you become aware of a new explanation for the evidence at your disposal which you had not considered before. This can happen in many fields of inquiry: medicine, law, science, everyday affairs. Here is a scenario by Steele & Stefánsson (2021):

> FRIENDS: 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, as opposed to their being warm friends or mere acquaintances. 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 with your friend, relative, for instance, to their being warm friends. (Steele & Stefánsson, 2021, sec. 5, Example 2)

Initially, the algebra contained the hypotheses 'my partner and my best friend met to have an affair' (*affair*) and 'my partner and my best friend met as friends' (*friends*). These were the only explanations you considered for the fact that your partner and your best friend met one night without telling you. At first, the hypothesis *affair* seems more likely than *friends*.[2] But, when the algebra changes, a new hypothesis is added which you had not considered before: your partner and your best friends met to plan a surprise party for your upcoming birthday (*surprise*). This is a better explanation. So, *surprise* now seems more likely than *affair*. And since *surprise* implies *friends*, the latter must be more likely than *affair*. This conclusion violates Reverse Bayesianism since the ratio of the probabilities of *friends* and *affair* has changed before and after awareness expansion.

Steele & Stefánsson note that a quick fix is available. It is reasonable to suppose that no

[2] This assumes that the prior probabilities of the two hypotheses were not strongly skewed in one direction. If you were initially nearly certain your partner could not possibly have an affair, even the fact they behaved very secretively or lied to you might not affect the probability of the two hypotheses.

change in the probabilities should occur so long as we confine ourselves to the old probability space. With this in mind, consider the following condition, called **Awareness Rigidity**:

$$\mathsf{P}^+(A|T^*) = \mathsf{P}(A),$$

where $T^*$ corresponds to a proposition that picks out, from the vantage point of the new awareness state, the entire possibility space *before* the episode of awareness growth. Awareness rigidity establishes that, once a suitable proposition $T^*$ is identified, the old probability assignments remain unchanged conditional on $T^*$. In our running example, ¬*Surprise* is the suitable proposition $T^*$: *that* there were was no surprise party in the making picks out the original possibility space. Conditional on ¬*Surprise*, no probability assignment should change, including the probability of *Affair*. This is the intended result.

But this is not the end of the story. Steele & Stefánsson go on to show that Awareness Rigidity does not hold in other cases, what they call *awareness refinement*. These are cases in which (roughly) the new proposition added to the algebra induces a more fine-grained partition of the possibility space. Consider this scenario:

> 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 & Stefánsson, 2021, sec. 5, Example 3)

You initially categorized movies by just language and genre, and then you refined your categorization by adding another variable, level of difficulty. Without considering language difficulty, you assigned the same probability to the hypotheses *thriller* and *comedy*. But learning that the owner was simple-minded made you think that the level of linguistic difficulty must be low and the movie most likely a thriller rather than a comedy (perhaps because thrillers are simpler—linguistically—than comedies). Since the probability of *thriller* goes up, MOVIES violates (against Reverse Bayesianism) the condition $\frac{\mathsf{P}(thriller)}{\mathsf{P}(comedy)} = \frac{\mathsf{P}^+(thriller)}{\mathsf{P}^+(Comedy)}$. For the same reason, it also violates (against Awareness Rigidity) the condition $\mathsf{P}(thriller) = \mathsf{P}^+(thriller|thriller \vee comedy)$, where $thriller \vee comedy$ is a proposition that picks out the entire possibility space.[3]

Some might object that the probability of *thriller* goes up, not because of awareness refinement, but because you learn that the owner is simple-minded. And if learning in the strict Bayesian sense—one modeled by conditionalization—takes place, it is should be no surprise that probabilities will shift. We will see, however, cases of awareness refinement that do no involve learning in the Bayesian sense and still violate Reverse Bayesianism and Awareness Rigidity. So it is incumbent to understand under what circumstances these principles fail.

As will become clear, we believe that theorizing about awareness growth should be grounded in the subject-matter information underlying the scenario at hand. This subject-matter takes many forms. In FRIENDS, awareness expansion does not change the basic presupposition that someone's behavior must have a reason. In MOVIES, awareness refinement does not change the fact that certain characteristics, such as language, difficulty or genre, may influence one's decision to select a movie for showing rather than another. Arguably, what is wrong with principles such as Reverse Bayesianism or Awareness Rigidity is that they are purely formal. In contrast, we need a formalism that can—at least in part—represent the relevant subject-information. In what follows, we illustrate how Bayesian networks can be of service.

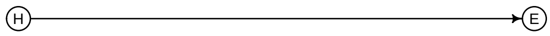
## 3 Expansion with Bayesian Networks

A Bayesian network is a compact formalism to represent probabilistic dependencies. It consists of a direct acyclic graph (DAG) accompanied by a probability distribution. The nodes in the graph represent random variables that can take different values. We will use 'nodes' and 'variables' interchangeably. The nodes are connected by arrows, but no loops are allowed, hence the name direct acyclic graph. Bayesian networks are relied upon in many fields, but have been rarely deployed to model awareness growth (the exception is Williamson (2003)). We think instead they are a good framework for this purpose. Awareness growth can be modeled as

[3] Since MOVIES is a case of refinement, $thriller \vee comedy$ picks out the entire possibility space both before and after awareness growth.

a change in the graphical network—nodes and arrows are changed, added or erased—as well as a change in the probability distribution from the old to the new network. In this section, we focus on cases of awareness expansion and turn to refinement in the next section.

Recall the scenario FRIENDS from before. To model it using a Bayesian network, we start with this graph, which is the usual hypothesis-evidence idiom (Fenton, Neil, & Lagnado, 2013):




where $H$ is the hypothesis node and $E$ the evidence node. If an arrow goes from $H$ to $E$, the full probability distribution associated with the Bayesian network is defined by two probability tables.[4] One table defines the prior probabilities $\mathsf{P}(H = h)$ for all the states (or values) of $H$, and another table defines the conditional probabilities of the form $\mathsf{P}(E = e|H = h)$, where uppercase letters represent the variables (nodes) and lower case letters represent the states (or values) of these variables. These two probability tables are sufficient to specify the full probability distribution. The other probabilities—say $\mathsf{P}(E = e)$, $\mathsf{P}(H = h|E = e)$, etc.—follow by simply applying the probability axioms.

Initially, before awareness growth, the hypothesis node $H$ takes only two states, *friends/acquaintances* and *affair*. These two states are meant to be exhaustive, so *affair* functions as the negation of *friends/acquaintances*, and vice versa. After awareness growth—in fact, awareness expansion—the two states are no longer exhaustive. Node $H$ now has a third state: *surprise*. So, expansion simply consists in the addition of an extra state in one of the nodes of the network. The rest of the structure of the network remains intact.

In FRIENDS, it is plausible to assume that the novel state added to the upstream node $H$ does not change the relative plausibility to the two hypotheses initially under consideration. Even after awareness expansion, the evidence *secretive* makes better sense in light of the hypothesis *affair* compared to *friends/acquaintances*, despite the fact that the novel hypothesis *surprise* makes better sense of the evidence overall. Thus, for all values $e$ and $h$ of upstream node $H$ and downstream node $E$ in the old network, the following constraint holds:

$$\frac{\mathsf{P}(E = e|H = h)}{\mathsf{P}(E = e|H = h)} = \frac{\mathsf{P}^{+}(E = e|H = h \,\&\, X \neq x^{*})}{\mathsf{P}^{+}(E = e|H = h \,\&\, X \neq x^{*})}, \tag{C}$$

[4]A major point of contention in the interpretation of Bayesian networks is is the meaning of the directed arrows. They could be interpreted causally—as though the direction of causality proceeds from the events described by the hypothesis to event described by the evidence—but they need not be; see footnote 10.

where $x^*$ is the new state added and $X$ is the node (upstream or downstream) to which the new state belongs, such as $H = surprise$ in FRIENDS. Constraint (C) is a restricted variant of Reverse Bayesianism that only applies to the conditional probabilities in the probability table for Bayesian networks of the form $H \to E$. The constraint mimics to some extent Awareness Rigidity in that it ensures that the conditional probabilities exclude the novel state $X = x^*$.

To gain a firm grasp on constraint (C), we shall examine a couple of examples by Mathani (2020). In her reading, these examples are meant to challenge the traditional distinction between expansion and refinement, and serve as counterexamples to Reverse Bayesianism. When modeled with Bayesian networks, these straightforward cases of awareness expansion and do violate constraint (C). The first of Matahani's examples goes like this:

> TENANT: 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 ... *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 ... Now let's suppose the possibility suddenly occurs to you that there might be another tenant living in the same flat (*Other*).

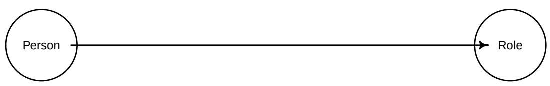
Initially, you thought the singer could either be the landlord or Bob, the tenant. Then you come to the realization that a third person could be the singer, another tenant. This scenario is a bit more complicated than FRIENDS. For one thing, the possibility that there could be a third person in the shower—besides Bob or the landlord—is a novel explanation for why you hear singing in the shower. So TENANT seems to be a standard case of expansion. The expansion in awareness, however, goes along with an interesting conceptual shift. Before awareness expansion, that Bob is in the shower and that a tenant is in the shower are equivalent descriptions, but after the expansion, this equivalence breaks down.

As Mathani shows, this scenario is problematic for Reverse Bayesianism. Suppose, after you hear singing in the shower, you become sure someone is in there, but you cannot tell who. So $\mathsf{P}(Landlord) = \mathsf{P}(Bob) = 1/2$, and since *Bob* and *Tenant* are equivalent, also $\mathsf{P}(Tenant)$ = 1/2. Now, *Landlord*, *Bob* and *Tenant* are all propositions that you were originally aware of, and thus Reverse Bayesianisn requires that their probabilities should remain in the same proportion after your awareness grows. But note that *Other* entails *Tenant* and *Bob* and *Other* are disjoint, so it follows that $\mathsf{P}^+(Other)$ must have zero probability. If $\mathsf{P}^+(Other)$ were greater than zero, the proportion of of the probability of *Tenant* to *Landlord* (or the proportion of the probability of *Bob* to *Landlord*) should change.

Here is perhaps a more intuitive way to see the challenge. In TENANT, it is natural to assign $1/3$ to *Landlord*, *Bob* and *Other* after awareness growth. That someone is singing in the shower is evidence that someone must be in there, but without any more discriminating evidence, each person should be assigned the same probability. Consequently, a probability of 2/3 should be assigned to *Tenant*. On this picture, the proportion of *Landlord* to *Tenant* changes from 1:1 (before awareness growth) to 1:2 (after awareness growth). This violates Reverse Bayesianism.

It is less clear whether this scenario is a challenge for Awareness Rigidity. Much depend on the choice of the proposition $T^*$ that picks, from the vantage point of the new awareness state, the old possibility space prior to awareness growth. The proposition $Landlord \vee Tenant$ picks out the entire possibility space, but yields the wrong results. For $\mathsf{P}^+(Landlord|Landlord \vee Tenant)$ and $\mathsf{P}^+(Bob|Landlord \vee Tenant)$ should both equal $1/2$, thus forcing $\mathsf{P}^+(Other|Landlord \vee Tenant)$ to zero. But, arguably, $Landlord \vee Tenant$ picks out a possibility space larger than the old one, becaue it also includes the possibility that another person would be in the shower. So, the right proportion $T^*$ is something like $Landlord \vee Bob$. This does the job. For $\mathsf{P}^+(Landlord|Landlord \vee Bob)$ and $\mathsf{P}^+(Bob|Landlord \vee Bob)$ should both equal $1/2$, and thus $\mathsf{P}^+(Other|Landlord \vee Bob)=0\$$, but this does not mean that $\mathsf{P}^+(Other|Landlord \vee Tenant)$ should equal zero. This is the intended result.

Bayesian networks can help to model this scenario. We start with the following graph:




Initially, the upstream node *Person* has two possible states, representing who is in the bathroom singing: *landlord-person* and *bob*. To simplify things, the assumption here is that the evidence of singing has already ruled out the possibility that no one would be in the shower. The downstream node *Role* has also two values, *landlord* and *tenant*. After your awareness grows, the upstream node *Person* should now have one more possible state, *other*.

The scenarios FRIENDS and TENANT are structurally identical as far as their modeling using Bayesian networks. So, if constraint (C) holds in FRIENDS, as seen earlier, it must also hold in TENANT. This is precisely what happens. For these ratios remain fixed:

$$\frac{\mathsf{P}(Role = landlord|Person = landlord)}{\mathsf{P}(Role = landlord|Person = bob)} = \frac{\mathsf{P}^{+}(Role = landlord|Person = landlord)}{\mathsf{P}^{+}(Role = landlord|Person = bob)} = 1/0$$

$$\frac{\mathsf{P}(Role = tenant|Person = landlord)}{\mathsf{P}(Role = tenant|Person = bob)} = \frac{\mathsf{P}^{+}(Role = tenant|Person = landlord)}{\mathsf{P}^{+}(Role = tenant|Person = bob)} = 0/1$$

So constraint (C) seems a plausible principle for governing cases of awareness expansion. The definition of awareness expansion we are working with is quite simple: whenever a new state is added to one of te node in the network, awareness expansion takes place. The novel state can be added to any node in the network and this would sufficient to quality for awareness expansion according to this definition. Intuitively the idea is that each node, with its range of states, characterized a possible exhaustive partition of the possibility space. each state is exclusive and their disjunction is exhaustive. Whenever a new state is added to a node, that existing partition is shown to be inadequate. Refinement, as we will see in the next section, simply consists in the addition of a new node without the need to reconfigure an existing partition.

The upshot of this discussion is this. Awareness expansion can be modeled by changes in the Bayesian network used to represent the epistemic state of the agent. The structure of the network does not change, but values are added to one of the nodes.

(We will consider changes to the downstream node shortly.) This addition can be carried while satisfying a restricted version of Reverse Bayesian, what we called constraint (C). This constraint outperforms both Reverse Bayesianism (which fails in FRIENDS) and Awareness Rigidity (which fails in TENANT).

What would happen is the extra value is added to the downstream node in the hypothesis-evidence network? Consider a variation of FRIENDS. Suppose that the evidence node $E$ could initially take only two values, say *Secretive* and *Public*. You then realize the evidence node could also take a third value, say *Ambigous*. This realization mandate a change in the old conditional probabilities $\mathsf{P}(E = e|H = h)$. Since *Secretive* and *Public* were initially considered exhaustive. Since the change is downstream the old conditional probabilities $\mathsf{P}(E = e|H = h)$ would remain the same, and thus *a fortiori*, constraint (C) would be satisfied.

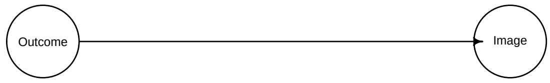
Consider now Mathani's second counterexample:

> COIN: You know that I am holding a fair ten pence UK coin which I am about to toss. You have a credence of 0.5 that it will land *Heads*, and a credence of 0.5 that it will land *Tails*. You think that the tails side always shows an engraving of a lion. So you also have a credence of 0.5 that (*Lion*) it will land with the lion engraving face-up: relative to your state of awareness *Tails* and *Lion* are equivalent.... Now let's suppose that you somehow become aware that occasionally ten pence coins have .... an engraving of Stonehenge on the tails side.

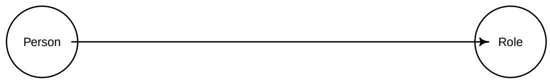
*Tails* and *Lion* are equivalent propositions prior to awareness growth. Suppose you initially gave *Tails* and *Lion* the same credence. Reverse Bayesianism requires that their relative proportions should stay the same after awareness grow. The same applies to *Heads* and *Tails*. But since *Lion* and *Stonehenge* are incompatible and the latter entails *Tails*, you should have $\mathsf{P}^{+}(\textit{Stonehenge}) = 0$, again an undesirable conclusion.

Mathani notes that COIN has the same structure as TENANT. This is true to some extent, but there is also an interesting asymmetry between the two scenarios. In TENANT, it is natural to assign $1/3$ to *Landlord*, *Bob* and *Other* after awareness growth. That someone is singing in the shower is evidence that someone must be in there, but without any more discriminating evidence, each person should be assigned the same probability. Consequently, a probability of 2/3 should be assigned to *Tenant*. On this picture, the proportion of *Landlord* to *Tenant* changes from 1:1 (before awareness growth) to 1:2 (after awareness growth). But, in COIN, the relative proportion of *Heads* to *Tails* should remain constant throughout, unless evidence emerges that the coin is not fair. One might have expected that *Landlord* and *Tenant* would behave just like *Heads* and *Tails*, but actually they do not.

Bayesian networks can help to model the asymmetry between these two scenarios. Consider COIN first. The structure of the scenario is represented by the following graph:




The upstream node *Outcome* has two states, *tails* and *heads*. These two states remain the same throughout. What changes are the states associated with the *Imagine* node downstream. Before awareness growth, the node *Image* has two states: *lions* and *heads-image*.[5] You assume that *Image* $=$ *lions* is true if and only if *Outcome* $=$ *tails* is true. Then, you come to the realization that the imagines for tails include a lion or a stonehenge engraving. So, after awareness growth, the node *Image* contains three states: *lion*, *stonehenge* and *heads-image*. Consider now the other scenario, TENANT. We start with the following graph:




Initially, the upstream node *Person* has two possible states, representing who is in the bathroom singing: *landlord-person* and *bob*. To simplify things, the assumption here is that the evidence of singing has already ruled out the possibility that no one would be in the shower. The downstream node *Role* has also two values, *landlord* and *tenant*. After your awareness grows, the upstream node *Person* should now have one more possible state, *other*.

The difference in modeling the two scenarios is this. In COIN, the states of the upstream node remain fixed, whereas in TENANT, they change. After awareness growth, no new state is added to *Outcome*, but an additional state, *other*, is added to *Person*. Plausible probability distributions for the Bayesian networks associated with the two scenarios are displayed in Table 1. How the networks should be built and which probabilities should shift is based on our background knowledge. This knowledge tells us that the equiprobability of *heads* and *tails* should not be affected by realizing that *stonhenge* is another possible engraving for the tails side. It also tells us that the probabilities of *landlord* and *tenant* should be affected by realizing that a third person could be in the shower.

The challenge now is to develop a systematic method to determine when constraint (C) is satisfied and when it fails. The structure of the Bayesian network will be our guide. This will afford us a firmer foundation to develop a general theory of awareness growth.

[5]The heads side must have some image, not specified in the scenario.

| $\mathsf{P}(\mathit{Image}\mid\mathit{Outcome})$ | | *Outcome* | |
|---|---|---|---|
| | | *heads* | *tails* |
| *Image* | *lion* | 0 | 1 |
| | *heads-image* | 1 | 0 |

| $\mathsf{P}^{+}(\mathit{Image}\mid\mathit{Outcome})$ | | *Outcome* | |
|---|---|---|---|
| | | *heads* | *tails* |
| *Image* | *lion* | 0 | 1/2 |
| | *stonehenge* | 0 | 1/2 |
| | *heads-image* | 1 | 0 |

| $\mathsf{P}(\mathit{Outcome}) = \mathsf{P}^{+}(\mathit{Outcome})$ | *Outcome* | |
|---|---|---|
| | *heads* | *tails* |
| | 1/2 | 1/2 |

| $\mathsf{P}(\mathit{Role}\mid\mathit{Person})$ | | *Person* | |
|---|---|---|---|
| | | *landlord-person* | *bob* |
| *Role* | *tenant* | 0 | 1 |
| | *landlord* | 1 | 0 |

| $\mathsf{P}^{+}(\mathit{Role}\mid\mathit{Person})$ | | *Person* | | |
|---|---|---|---|---|
| | | *landlord-person* | *bob* | *other* |
| *Role* | *tenant* | 0 | 1/2 | 1/2 |
| | *landlord* | 1 | 0 | 0 |

| $\mathsf{P}(\mathit{Person})$ | *Person* | |
|---|---|---|
| | *landlord-person* | *bob* |
| | 1/2 | 1/2 |

| $\mathsf{P}^{+}(\mathit{Person})$ | *Person* | | |
|---|---|---|---|
| | *landlord-person* | *bob* | *other* |
| | 1/3 | 1/3 | 1/3 |

Table 1: Top table displays a plausible probability distribution for COIN and bottom table does the same for TENANT.

## 4 Refinement

We turn now from cases of awareness expansion to cases of awareness refinement. In the framework of Bayesian networks, expansion consisted in added values to nodes in the network. Refinement, instead, can be modeled by adding nodes to the network without necessarily add any new values to the existing nodes. Intuitively, refinement takes place when an epistemic agent acquires a more-fined grained picture of the situation, say instead of thinking that the political spectrum is divided into liberal and conservatives, the political spectrum can be further divided into traditional-liberal, new-liberal, traditional-conservative and new-conservative. The political spectrum is still divided into liberal and conservative—not expansion occurred—but these two categories have been further refined.

Although there is no shortage of counterexamples to Reverse Bayesianism when it comes to awareness refinement, we begin with our own. This will allow us to underscore the role of subject-matter assumptions in theorizing about awareness growth. Consider the following scenario:

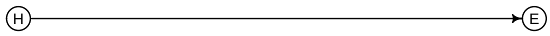
> LIGHTING: You have evidence that favors a certain hypothesis, say a witness saw the defendant around the crime scene. You give some weight to this evidence. In your assessment, that the defendant was seen around the crime scene (your evidence) raises the probability that the defendant was actually there (your hypothesis). But now you ask, what if it was dark when the witness saw the defendant? In light of your realization that it could have been dark, you wonder whether (and if so how) you should change the probability that you assigned to the hypothesis that the defendant was around the crime scene.

As your awareness grows, you do not learn anything specific about the lighting conditions, neither that they were bad nor that they were good. You simply wonder what they were, a variable you had previously not considered. So no Bayesian updating takes place in the strict sense, although broadly speaking some new information has been introduced.[6] Something has

[6] HERE EXPLAIN DIFFERENCE WITH STEELE AND STEFANSSON. The process of awareness growth in LIGHTING adds only one extra variable, lighting conditions, while MOVIES adds two extra variables, language difficulty and whether the owner is simple-minded or not. Further, MOVIES contains a clear-cut case of learning, that the owner *is* simple-minded. This is not so in LIGHTING. Strictly speaking, you are learning that it is *possible* that the lighting conditions were bad. However, you are not conditioning on the proposition 'the lighting conditions were bad' or 'the lighting conditions were good'. So you are not learning about the lighting conditions in the sense in which learning is understood in this paper.

changed in your epistemic state—you have a more fine-grained assessment of what could have happened—but it is not clear what you should do in this scenario. Since the lighting conditions could have been bad but could also have been good, perhaps you should just stay put until you learn something more specific.

In what follows, we illustrate how Bayesian networks helps to model what is going on in LIGHTING and conclude that you should probably revise downward your confidence in the hypothesis that the defendant was around the crime scene. The starting point of our analysis is the usual hypothesis-evidence idiom, repeated below for convenience:

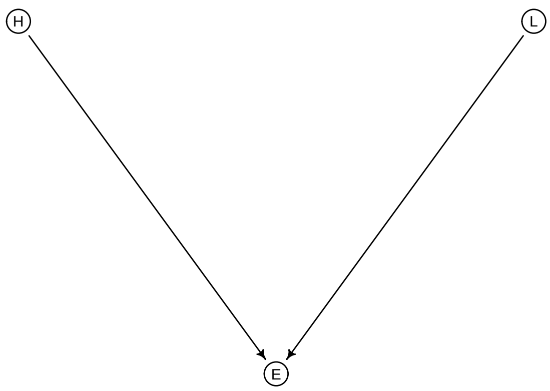



Since you trust the evidence, you think that the evidence is more likely under the hypothesis that the defendant was present at the crime scene than under the alternative hypothesis:

$$\mathsf{P}(E{=}\mathit{seen}|H{=}\mathit{present}) > \mathsf{P}(E{=}\mathit{seen}|H{=}\mathit{absent})$$

The inequality is a qualitative ordering of how plausible the evidence is in light of competing hypotheses. No matter the numbers, by the probability calculus, it follows that the evidence raises the probability of the hypothesis *H=present*.

Now, as you wonder about the lighting conditions, the graph should be amended:




where the node *L* can have two values, *L=good* and *L=bad*. Commonsense as well as psychological findings suggest that when the visibility deteriorates, people's ability to identify faces worsen. So a plausible way to modify your assessment of the evidence is as follows:

$$\mathsf{P}^{+}(E{=}\mathit{seen}|H{=}\mathit{present} \wedge L{=}\mathit{good}) > \mathsf{P}^{+}(E{=}\mathit{seen}|H{=}\mathit{absent} \wedge L{=}\mathit{good})$$

$$\mathsf{P}^{+}(E{=}seen|H{=}present \wedge bad) = \mathsf{P}^{+}(E{=}seen|H{=}absent \wedge L{=}bad)$$

In words, if the lighting conditions were good, you still trust the evidence like you did before (first line), but if the lighting conditions were bad, you regard the evidence as no better than chance (second line). These probabilistic constraints are plausible, but should ultimately be grounded on verifiable empirical regularities.

Despite the change in awareness, you have not learned anything in the strict sense. Your new stock of evidence does not contain neither the information that the lighting conditions were bad nor that they were good. But the Bayesian network structure that represents your epistemic state is now more fine-grained. The network contains the new variable $L$ which it did not contain prior to the episode of awareness growth. In addition—and this is the crucial point—the new variable bears certain *structural relationships* with the variables $H$ and $E$. The graphical network represents a direct probabilistic dependency between the lighting conditions $L$ and the witness sensory experience $E$, but does not allow for any direct dependency between the lighting conditions and the fact that the defendant was (or was not) at the crime scene. There is no direct arrow between the nodes $L$ and $H$. This structure of dependencies captures our causal intuitions about the scenario: the lighting conditions do affect what the witness could see, but do not directly affect what the defendant might have have done.

Without Bayesian networks, episodes of awareness growth could only be modeled by the addition of new propositions that were not previously in the algebra. But this approach would fail to capture crucial information. When awareness growth takes place against the background of an intuitive causal structure of the world—as in the case of LIGHTING—this structure should also be modeled. Bayesian networks offer a formal framework that can do precisely that.

This model of causal structure can now guide us to decide whether the restricted version of Reverse Bayesianism, what we called constraint (C), holds in this scenario, Specifically, we need to assess whether the following holds:

$$\frac{\mathsf{P}(E = seen|H = present)}{\mathsf{P}(E = seen|H = absent)} = \frac{\mathsf{P}^{+}(E = seen|H = present)}{\mathsf{P}^{+}(E = seen|H = absent)}.$$

The question here is whether you should assess the evidence at your disposal—that the witness saw the defendant at the crime scene—any differently than before. As noted earlier, without a clear model of the scenario, it might seem that you should simply stay put. After all, besides the sensory experience of the witness, you have gained no novel information about the lighting conditions. Should you thus conclude that the evidence has the same value before and after the realization that lighting could have been bad?

The evidence would have the same value if the likelihood ratios associated with it relative to the competing hypotheses were the same before and after awareness growth. But, in changing the probability function from $\mathsf{P}()$ to $\mathsf{P}^+()$, it would be quite a coincidence if this were true. In our example, many possible probability assignments violate this equality. If before awareness growth you thought the evidence favored the hypothesis *H=present* to some extent, after the growth in awareness, the evidence is likely to appear less strong.[7] If this is correct, this outcome violates constraint (C). Reverse Bayesianism is also violated since the ratio of the probabilities

[7] By the law of total probability, the right hand side of the equality in (C) should be expanded, as follows:

$$\frac{\mathsf{P}^+(E=e|H=h)}{\mathsf{P}^+(E=e|H=h')} = \frac{\mathsf{P}^+(E{=}seen \wedge L{=}good|H{=}present) + \mathsf{P}^+(E{=}seen \wedge L{=}bad|H{=}present)}{\mathsf{P}^+(E{=}seen \wedge L{=}good|H{=}absent) + \mathsf{P}^+(E{=}seen \wedge L{=}bad|H{=}absent)}.$$

For concreteness, let's use some numbers:

$$\mathsf{P}(E{=}seen|H{=}present) = \mathsf{P}^+(E{=}seen|H{=}present \wedge L{=}good) = .8$$
$$\mathsf{P}(E{=}seen|H{=}absent) = \mathsf{P}^+(E{=}seen|H{=}absent \wedge L{=}good) = .4$$
$$\mathsf{P}^+(E{=}seen|H{=}present \wedge L{=}bad) = \mathsf{P}^+(E{=}seen|H{=}absent \wedge L{=}bad) = .5.$$
$$\mathsf{P}^+(L{=}bad) = \mathsf{P}^+(L{=}good) = .5.$$

So the ratio $\frac{\mathsf{P}(E=seen|H=present)}{\mathsf{P}(E=seen|H=absent)}$ equals 2. After the growth in awareness, the ratio $\frac{\mathsf{P}^+(E=seen|H=present)}{\mathsf{P}^+(E=seen|H=absent)}$ will drop to $\frac{.65}{.45} \approx 1.44$. The calculations here rely on the dependency structure encoded in the Bayesian network (see starred step below).

$$\begin{aligned}
\mathsf{P}^+(E{=}seen|H{=}present) &= \mathsf{P}^+(E{=}seen \wedge L{=}good|H{=}present) + \mathsf{P}^+(E{=}seen \wedge L{=}bad|H{=}present)\\
&= \mathsf{P}^+(E{=}seen|H{=}present \wedge L{=}good) \times \mathsf{P}^+(L{=}good|H{=}present)\\
&+ \mathsf{P}^+(E{=}seen|H{=}present \wedge L{=}bad) \times \mathsf{P}^+(L{=}bad|H{=}present)\\
&=^* \mathsf{P}^+(E{=}seen|H{=}present \wedge L{=}good) \times \mathsf{P}^+(L{=}good)\\
&+ \mathsf{P}^+(E{=}seen|H{=}present \wedge L{=}bad) \times \mathsf{P}^+(L{=}bad)\\
&= .8 \times .5 + .5 * .5 = .65
\end{aligned}$$

$$\begin{aligned}
\mathsf{P}^+(E{=}seen|H{=}absent) &= \mathsf{P}^+(E{=}seen \wedge L{=}good|H{=}absent) + \mathsf{P}^+(E{=}seen \wedge L{=}bad|H{=}absent)\\
&= \mathsf{P}^+(E{=}seen|H{=}absent \wedge L{=}good) \times \mathsf{P}^+(L{=}good|H{=}absent)\\
&+ \mathsf{P}^+(E{=}seen|H{=}absent \wedge L{=}bad) \times \mathsf{P}^+(L{=}bad|H{=}absent)\\
&=^* \mathsf{P}^+(E{=}seen|H{=}absent \wedge L{=}good) \times \mathsf{P}^+(L{=}good)\\
&+ \mathsf{P}^+(E{=}seen|H{=}absent \wedge L{=}bad) \times \mathsf{P}^+(L{=}bad)\\
&= .4 \times .5 + .5 * .5 = .45
\end{aligned}$$

This argument can be repeated with many other numerical assignments.

of *H=present* to *E=seen*, before and after awareness growth, has changed:

$$\frac{\mathsf{P}^{E=seen}(H=present)}{\mathsf{P}^{E=seen}(E=seen)} \neq \frac{\mathsf{P}^{+,E=seen}(H=present)}{\mathsf{P}^{+,E=seen}(E=seen)},$$

where $\mathsf{P}^{E=seen}()$ and $\mathsf{P}^{+,E=seen}()$ represent the agent's degrees of belief, before and after awareness growth, updated by the evidence *E=seen*.[8]

The general lesson to be learned here has to do with the importance of formalizing structural assumptions and the role of Bayesian networks in modeling awareness growth. Modeling those structural assumptions allows us to see that constraint (C)—as well as Reverse Bayesianism more generally—fails here. To strengthen this point, consider this variation of the LIGHTING scenario:

> VERACITY: A witness saw that the defendant was around the crime scene and you initially took this to be evidence that the defendant was actually there. But then you worry that the witness might be lying or misremembering what happened. Perhaps, the witness was never there, made things up or mixed things up. Should you reassess the evidence at your disposal? If so, how?

It might seem that this scenario is no different from LIGHTING. The realization that lighting could be bad should make you less confident in the truthfulness of the sensory evidence. And the same conclusion should presumably follow from the realization that the witness could be lying. So both scenarios would be counterexamples to Reverse Bayesianism. But, upon closer scrutiny, things are not that simple. To run the two scenarios together would be a mistake.

The evidence at your disposal in LIGHTING is the sensory evidence—the experience of seeing—and the possibility of bad lighting does affect the quality of your visual experience. So, if lighting was indeed bad, this would warrant lowering your confidence in the truthfulness of the visual experience. But the possibility of lying in VERACITY does not affect the quality of the visual experience in and of itself, although it affects the quality of the *reporting* of that experience. So, if the witness did lie, this would not warrant lowering your confidence

[8]The scenario also violates Awareness Rigidity which requires that $\mathsf{P}^{+}(A|T^{*}) = \mathsf{P}(A)$, where $T^{*}$ corresponds to a proposition that picks out, from the vantage point of the new awareness state, the entire possibility space before the episode of awareness growth. In LIGHTING, however, $T^{*}$ does not change, so Awareness Rigidity would require that $\mathsf{P}^{+}(A) = \mathsf{P}(A)$, and instead in the scenario, we have

$$\mathsf{P}^{+}(H=present|E=seen) \neq \mathsf{P}(H=present|E=seen).$$

in the truthfulness of the visual experience, only in the truthfulness of the reporting of that experience. The distinction between the visual experience and its reporting is crucial here. Bayesian networks help to model this distinction precisely, and then see why LIGHTING and VERACITY are structurally different.

The graphical network should initially look like the initial DAG for LIGHTING, consisting of the hypothesis node $H$ upstream and the evidence node $E$ downstream. As your awareness grows, the graphical network should be updated by adding another node $R$ further downstream:

$$H \longrightarrow E \longrightarrow R$$

As before, the hypothesis node $H$ bears on the whereabouts of the defendant and has two values, $H$=*present* and $H$=*absent*. Note the difference between $E$ and $R$. The evidence node $E$ bears on the visual experience had by the witness. The reporting node $R$, instead, bears on what the witness reports to have seen. The chain of transmission from 'visual experience' to 'reporting' may fail for various reasons, such as lying or misremembering.

In VERACITY, the conditional probabilities, $\mathsf{P}(E = e|H = h)$ should be the same as $\mathsf{P}^{+}(E = e|H = h)$ for any values $e$ and $h$ of the variables $H$ and $E$ that are shared before and after awareness growth. In comparing the old and new Bayesian network, this equality falls out from their structure, as the connection between $H$ and $E$ remains unchanged. Thus, constraint (C)—along with Reverse Bayesianism—is perfectly fine in scenarios such as VERACITY.

This does not mean that the assessment of the probability of the hypothesis $H$=*present* should undergo no change. If you worry that the witness could have lied, this should presumably make you less confident about $H$=*present*. To accommodate this intuition, VERACITY can be interpreted as a scenario in which an episode of awareness refinement takes place together with a form of retraction. At first, after the learning episode, you update your belief based on the *visual experience* of the witness. But after the growth in awareness, you realize that your learning is in fact limited to what the witness *reported* to have seen. The previous learning episode is retracted and replaced by a more careful statement of what you learned: instead of conditioning on $E$=*seen*, you should condition on what the witness reported to have seen, $R$=*seen-reported*. This retraction will affect the probability of the hypothesis $H$=*present*.

Where does this leave us? Refinement cases that might at first appear similar can be structurally different in important ways, and this difference can be appreciated by looking at the Bayesian networks used to model them. In modeling VERACITY, the new node is added downstream, while in modeling LIGHTING, it is added upstream. This difference affects how probability assignments should be revised. Since the conditional probabilities associated with the upstream nodes are unaffected, Reverse Bayesianism is satisfied in VERACITY.[9] By contrast, since the conditional probabilities associated with the downstream node will often have to change, Reverse Bayesianism fails in LIGHTING.

This further corroborates our working hypothesis: structural features about how we conceptualize a specific scenario are the guiding principles about how we update the probability function through awareness growth, not a formal principle like Reverse Bayesianism. We further elaborate on this conjecture by drawing on some examples from Anna Mathani.

### 4.1 Sure no-gain bets

Suppose the witness reports to have seeing the defendant around the crime scene. You are not aware that the witness could be lying. Thus, you are 100% confident that the witness saw is what they report to have seeing. In fact, you make no distinction between reporting to have seeing and seeing itself. So you would be willing to buy for 1$ the following bet: if the witness saw the defendant, you get 1$, and 0$ otherwise. If the witness did see the defendant, you get you 1$ back, and otherwise you loose \1$. You are 100% sure the witness did see the defendant, so—by your lights—you stand to loose no money whatsoever from this bet. But suppose that, as a matter of fact, there is a difference between reporting and seeing. So,the witness might report to have seeing something without actually having seeing it. So, contrary to your conviction, that the witness saw the defendant is not 100% probable. This means that you would be willing to engage in a bet in which you are guaranteed not to win any money and could potentially lose money. If the witness did see the defendant you would get your 1$ back, but if not, you would lose it.

## 5 Towards a general theory

We conclude with some programmatic remarks. We think that the awareness of agents grows while holding fixed certain material structural assumptions, based on commonsense, semantic

[9] Note that $\mathsf{P}(H{=}\textit{present}|E{=}\textit{seen}) \neq \mathsf{P}(H{=}\textit{present}|R{=}\textit{seen-reported})$, but since you are conditioning on different propositions, this does not conflict with Reverse Bayesianism.

stipulations or causal dependency.[10] To model awareness growth, we need a formalism that can express these material structural assumptions. This can done using Bayesian networks, and we offered some illustrations of this strategy. These material assumptions also guide us in formulating the adequate conservative constraints, and these will inevitably vary on a case-by-case basis. The literature on awareness growth from a Bayesian perspective is primarily concerned with a formal, almost algorithmic solution to the problem. Insofar as Reverse Bayesianism is an expression of this formalistic aspiration, we agree with Steele and Stefánsson that we are better off looking elsewhere.

Awareness growth can occur in different ways. The key question is to what extent probability assignments that were made prior to the episode of awareness growth can be retained. There seems to no clear rule that can decide that. We propose the following procedure. Construct a Bayesian network prior to awareness growth and compare it with the new Bayesian network after awareness growth. If the new arrows and nodes are all downstream, the old probabilities table should not be changed. The paradigmatic cases of this are scenarios VERACITY and COIN. If, instead, the new arrows and and nodes are upstream, the old probabilities tables should be changed. The paradigmatic examples are LIGHTING and TENANT.

## References


Bradley, R. (2017). *Decision theory with a human face*. Cambridge University Press.

Chihara, C. S. (1987). Some problems for bayesian confirmation theory. *British Journal for the Philosophy of Science*, *38*(4), 551–560.

Earman, J. (1992). *Bayes or bust? A critical examination of bayesian confirmation theory*. MIT press.

Fenton, N., Neil, M., & Lagnado, D. A. (2013). A General Structure for Legal Arguments About Evidence Using Bayesian Networks. *Cognitive Science*, *37*(1), 61–102. https://doi.org/10.1111/cogs.12004

Glymour, C. (1980). *Theory and evidence*. Princeton University Press.

Howson, C. (1976). The development of logical probability. In *Essays in memory of imre lakatos. Boston studies in the philosophy of science* (pp. 277–298). Springer.

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

Lakatos, I. (1968). Changes in the problem of inductive logic. *Studies in Logic and the Foundations of Mathematics*, *51*, 315–417.

Mathani, A. (2020). Awareness growth and dispositional attitudes. *Synthese*, *198*(9), 8981–8997.


[10] Arrows in Bayesian networks are often taken to represent causal relationships, but other interpretations exist. Schaffer (2016) discusses an interpretation in which arrows represent grounding relations rather than causality.

Pettigrew, R. (forthcoming). How should your beliefs change when your awareness grows? *Episteme*, 1–25. https://doi.org/10.1017/epi.2022.33

Roussos, J. (2021). Awareness growth and belief revision. *Manuscript*.

Schaffer, J. (2016). Grounding in the image of causation. *Philosophical Studies*, *173*, 49–100.

Steele, K., & Stefánsson, O. (2021). Belief revision for growing awareness. *Mind*, *130*(520), 1207–1232.

Wenmackers, S., & Romeijn, J.-W. (2016). New theory about old evidence: A framework for open-minded bayesianism. *Synthese Volume*, *193*, 1225–1250.

Williamson, J. (2003). Bayesianism and language change. *Journal of Logic, Language, and Information*, *12*(1), 53–97.

Zabell, S. (1992). Predicting the unpredictable. *Synthese*, *90*(1), 205–232.