

Beyond Reverse Bayesianism:

Awareness Growth in Bayesian Networks

July 28, 2022

Wordcount (including footnotes): 5,392

We examine Steele and Stefánsson’s case against Reverse Bayesianism, a popular theory that addresses the problem of awareness growth. We show that Steele and Stefánsson’s counterexamples have limited applicability but agree with their skepticism toward Reverse Bayesianism. We strengthen their argument by providing a simpler counterexample that is less prone to objections. In addition, we submit that the problem of awareness growth cannot be tackled in an algorithmic manner, because subject-matter assumptions need to be made explicit. Thanks to their ability to express probabilistic dependencies, we sketch 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,

$$P^E(H) = P(H|E),$$

where $P()$ represents the agent’s old degree of belief (before the learning experience E) and $P^E()$ 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

1 degree of belief to any proposition conditional on E . This picture commits the agent to the
2 specification of their ‘total possible future experience’ (Howson, 1976), as though learning was
3 confined to an ‘initial prison’ (Lakatos, 1968).

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

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

17 Critics of Bayesianism and sympathizers alike have been discussing the problem of awareness
18 growth under different names for quite some time, at least since the eighties. This problem arises
19 in a number of different contexts, for example, new scientific theories (Chihara, 1987; Earman,
20 1992; Glymour, 1980), language changes and paradigm shifts (Williamson, 2003), and theories
21 of induction (Zabell, 1992). A proposal that has attracted considerable scholarly attention
22 in recent years is Reverse Bayesianism (Bradley, 2017; Karni & Vierø, 2015; Wenmackers
23 & Romeijn, 2016). The idea is to model awareness growth as a change in the algebra while
24 ensuring that the proportions of probabilities of the propositions shared between the old and
25 new algebra remain the same in a sense to be specified.

26 Let \mathcal{F} be the initial algebra of propositions and let \mathcal{F}^+ the algebra after the agent’s awareness
27 state has grown. Both algebras contain the contradictory and tautologous propositions \perp and
28 \top , and they are closed under connectives such as disjunction \vee , conjunction \wedge and negation \neg .
29 Denote by X and X^+ the subsets of these algebras that contain only basic propositions, namely

¹Roussos (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 distinction between true unawareness and temporary unawareness.

1 those without connectives. **Reverse Bayesianism** posits that the ratio of probabilities for any
2 basic propositions A and B in both X and X^+ —the basic propositions shared by the old and
3 new algebra—remain constant through the process of awareness growth:

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

4 where $P()$ represents the agent's degree of belief before awareness growth and $P^+()$ represents
5 the agent's degree of belief after awareness growth.

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

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

19 Here we offer a diagnosis of what is wrong with Reverse Bayesianism and outline an
20 alternative proposal. The problem of awareness growth—we hold—cannot be tackled in an
21 algorithmic manner because subject-matter assumptions, both probabilistic and structural, need
22 to be made explicit. So any theory of awareness growth cannot be a purely formal theory. This
23 does not mean, however, we should give up on probability theory altogether in this context.
24 Thanks to its ability to express probabilistic dependencies, we think that the theory of Bayesian
25 networks can help to model awareness growth in the Bayesian framework. We offer a number
26 of illustrations of this claim (§ 2).

27 We share Steele and Stefánsson's skepticism toward Reverse Bayesianism, but also believe
28 that their counterexamples have limited applicability (§ 5.1). To remedy that, we provide a

1 simpler counterexample that is less prone to objections (§ ??).

2 At the same time, we conjecture that the problem of awareness growth cannot be tackled in
3 an algorithmic manner because subject-matter assumptions, both probabilistic and structural,
4 need to be made explicit. Thanks to its ability to express probabilistic dependencies, we think
5 that the theory of Bayesian networks can help to model awareness growth in the Bayesian
6 framework. We offer two illustrations of this claim (§ 2).

7 **2 Structural assumptions**

8 Although there is no shortage of counterexamples to Reverser Bayesianism, we begin with our
9 own. This will allow us to underscore the role of structural assumptions in theorizing about
10 awareness growth. Consider the following scenario:

11 LIGHTING: You have evidence that favors a certain hypothesis, say a witness
12 saw the defendant around the crime scene. You give some weight to this evi-
13 dence. In your assessment, that the defendant was seen around the crime scene
14 (your evidence) raises the probability that the defendant was actually there (your
15 hypothesis). But now you wonder, what if it was dark when the witness saw
16 the defendant? Further, in light of your realization that it could have been dark,
17 you wonder whether (and if so how) you should change the probability that you
18 assigned to the hypothesis that the defendant was around the crime scene.

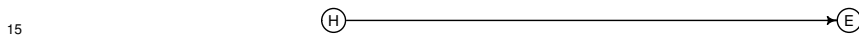
19 In this scenario, you do not learn anything specific, neither that the lighting conditions were
20 bad nor that they were good. You simply wonder about what they were. So no Bayesian
21 updating takes place in the strict sense, although broadly speaking some new information
22 has been introduced.² Something has changed in your epistemic state—you have a more
23 fine-grained assessment of what could have happened—but it is not clear what you should do in
24 this scenario. Since the lighting conditions could have been bad but could also have been good,
25 perhaps you should just stay put until you learn anything specific. In what follows, we illustrate

²HERE 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.

1 how Bayesian networks helps to model what is going on in LIGHTING and then show that you
2 should not stay put in this scenario and should probably revise downward your confidence in
3 the hypothesis that the defendant was around the crime scene.

4 A Bayesian network is compact formalism to represent probabilistic dependencies. A
5 Bayesian network consists of a direct acyclic graph (DAG) accompanied by a probability
6 distribution. The nodes in the graph represent random variables that can take different values.
7 We will use ‘nodes’ and ‘variables’ interchangeably. The nodes are connected by arrows, but
8 no loops are allowed, hence the name direct acyclic graph. Bayesian networks are relied upon
9 in many fields, but have been rarely deployed to model awareness growth (the exception is
10 Williamson (2003)). We think instead they are a good framework for this purpose. Awareness
11 growth can be modeled as a change in the graphical network—nodes and arrows are added or
12 erased—as well as a change in the probability distribution from the old to the new network.

13 To model LIGHTING with Bayesian networks, we start with this graph, which is the usual
14 hypothesis-evidence idiom:



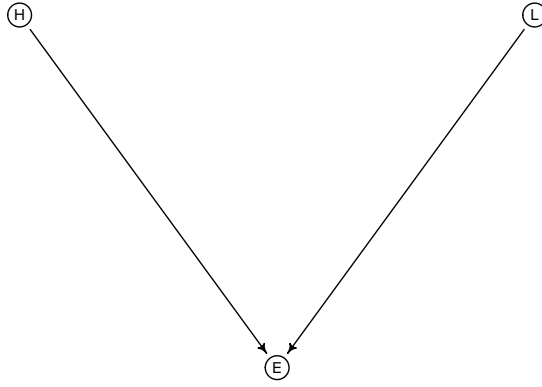
16 where H is the hypothesis node and E the evidence node. If an arrow goes from H to E , the
17 probability distribution associated with the Bayesian network should be defined by the prior
18 probabilities for all the states of H , and conditional probabilities of the form $P(E = e|H = h)$,
19 where uppercase letters represent the variables (nodes) and lower case letters represent the
20 values of these variables.³ Since you trust the evidence, you think that the evidence is more
21 likely under the hypothesis that the defendant was present at the crime scene than under the
22 alternative hypothesis:

$$P(E=seen|H=present) > P(E=seen|H=absent)$$

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

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

³A 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.



1

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

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

5

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

6 Here is what you are thinking: if the lighting conditions were good, you should still trust the
 7 evidence like you did before (first line). But if the lighting conditions were bad, you should
 8 regard the evidence as no better than chance (second line). These probabilistic constraints are
 9 plausible, but should ultimately be grounded in verifiable empirical regularities.

10 Despite the change in awareness, you have not learned anything definitive in the strict sense.
 11 Your new stock of evidence does not contain neither the information that the lighting conditions
 12 were bad nor that they were good. But the Bayesian network structure that represents your
 13 epistemic state is now more fine-grained. It contains the new variable L which it did not
 14 contain prior to the episode of awareness growth. In addition—and this is the crucial point—the
 15 new variable bears certain *structural relationships* with the variables H and E . The graphical
 16 network allows for a direct probabilistic dependency between the lighting conditions L and
 17 the witness sensory experience E , but does not allow for any direct dependency between
 18 the lighting conditions and the fact that the defendant was (or was not) at the crime scene.
 19 This structure of dependencies capture our causal intuitions about the scenario: the lighting
 20 conditions do affect what the witness could see, but do not directly affect what the defendant
 21 might have have done.

22 The general moral here is this: episodes of awareness growth should not simply be modeled

by the addition of new propositions that were not previously in the algebra. Since they take place against the background of an intuitive causal structure of the world, this structure should also be modeled. Bayesian networks offer a tried and tested formal framework that can do precisely that. (More on this soon.)

With a formal model of the scenario at hand, let us now tackle the question whether you should assess the evidence at your disposal—that the witness saw the defendant at the crime scene—any differently than before. 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:

$$\frac{P(E = e|H = h)}{P(E = e|H = h')} = \frac{P^+(E = e|H = h)}{P^+(E = e|H = h')}. \quad (C)$$

But, in changing the probability function from $P()$ to $P^+()$, it would be quite a coincidence if (C) 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.⁴

⁴By the law of total probability, the right hand side of the equality in (C) should be expanded, as follows:

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

For concreteness, let's use some numbers:

$$P(E=seen|H=present) = P^+(E=seen|H=present \wedge L=good) = .8$$

$$P(E=seen|H=absent) = P^+(E=seen|H=absent \wedge L=good) = .4$$

$$P^+(E=seen|H=present \wedge L=bad) = P^+(E=seen|H=absent \wedge L=bad) = .5.$$

$$P^+(L=bad) = P^+(L=good) = .5.$$

So the ratio $\frac{P(E=seen|H=present)}{P(E=seen|H=absent)}$ equals 2. After the growth in awareness, the ratio $\frac{P^+(E=seen|H=present)}{P^+(E=seen|H=absent)}$ 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).

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

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

1 Why does this matter? We have seen that, after awareness growth, you should typically
 2 regard the evidence $E=seen$ as one that favors $H=present$ less strongly. This outcome violates
 3 Reverse Bayesianism since the ratio of the probabilities of $H=present$ to $E=seen$, before and
 4 after awareness growth, has changed:

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

5 where $P^{E=seen}()$ and $P^{+,E=seen}()$ represent the agent's degrees of belief, before and after aware-
 6 ness growth, updated by the evidence $E=seen$.⁵

7 Besides counterexamples that can be leveled against Reverse Bayesianism, we think there
 8 is a more general lesson to be learned to which we have already alluded to. It has to do with
 9 the importance of formalizing structural assumptions and the role of Bayesian networks in
 10 modeling awareness growth. To strengthen this point, consider this variation of the LIGHTING
 11 scenario:

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

17 It might seem that this scenario is no different from LIGHTING. The realization that lighting
 18 could be bad should make you less confident in the truthfulness of the sensory evidence. And
 19 the same conclusion should presumably follow from the realization that the witness could be
 20 lying. But, upon closer scrutiny, things are not that simple. The evidence at your disposal
 21 in LIGHTING is the sensory evidence—the experience of seeing—and the possibility of bad
 22 lighting does warrant lowering your confidence in the truthfulness of the visual experience.
 23 But the possibility of lying in VERACITY does not warrant lowering your confidence in the

This argument can be repeated with many other numerical assignments.

⁵The scenario also violates a variant of Reverse Bayesianism, what (Steele & Stefánsson, 2021) call Awareness Rigidity. This requires that $P^+(A|T^*) = P(A)$, 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 $P^+(A) = P(A)$, and instead in the scenario, we have

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

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

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



9 As before, the hypothesis node H bears on the whereabouts of the defendant and has two values,
 10 $H=present$ and $H=absent$. Note the difference between E and R . The evidence node E bears
 11 on the visual experience had by the witness. The reporting node R , instead, bears on what the
 12 witness reports to have seen. The chain of transmission from ‘visual experience’ to ‘reporting’
 13 may fail for various reasons, such as lying or misremembering.

14 In VERACITY, the conditional probabilities, $P(E = e | H = h)$ should be the same as $P^+(E =$
 15 $e | H = h)$ for any values e and h of the variables H and E that are shared before and after
 16 awareness growth. In comparing the old and new Bayesian network, this equality falls out
 17 from their structure, as the connection between H and E remains unchanged. Thus, Reverse
 18 Bayesianism and Awareness Rigidity are perfectly fine in scenarios such as VERACITY.

19 This does not mean that the assessment of the probability of the hypothesis $H=present$ should
 20 undergo no change. If you worry that the witness could have lied, this should presumably
 21 make you less confident about $H=present$. To accommodate this intuition, VERACITY can
 22 be interpreted as a scenario in which an episode of awareness refinement takes place together
 23 with a form of retraction. At first, after the learning episode, you update your belief based on
 24 the *visual experience* of the witness. But after the growth in awareness, you realize that your
 25 learning is in fact limited to what the witness *reported* to have seen. The previous learning
 26 episode is retracted and replaced by a more careful statement of what you learned: instead
 27 of conditioning on $E=seen$, you should condition on what the witness reported to have seen,
 28 $R=seen-reported$. This retraction will affect the probability of the hypothesis $H=present$.

29 Where does this leave us? Refinement cases that might at first appear similar can be
 30 structurally different in important ways, and this difference can be appreciated by looking

1 at the Bayesian networks used to model them. In modeling VERACITY, the new node is
2 added downstream, while in modeling LIGHTING, it is added upstream. This difference affects
3 how probability assignments should be revised. Since the conditional probabilities associated
4 with the upstream nodes are unaffected, Reverse Bayesianism is satisfied in VERACITY.⁶ By
5 contrast, since the conditional probabilities associated with the downstream node will often
6 have to change, Reverse Bayesianism fails in LIGHTING.

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

11 3 Mathani's counterexamples

12 Mathani (2020) offers two counterexamples to Reverse Bayesianism. The first goes like this:

13 TENANT: Suppose that you are staying at Bob's flat which he shares with his
14 landlord. You know that Bob is a tenant, and that there is only one landlord, and
15 that this landlord also lives in the flat. In the morning you hear singing coming
16 from the shower room, and you try to work out from the sounds who the singer
17 could be. At this point you have two relevant propositions that you consider
18 possible ... *Landlord* standing for the possibility that the landlord is the singer,
19 and *Bob* standing for the possibility that Bob is the singer ... Because you know
20 that Bob is a tenant in the flat, you also have a credence in the proposition *Tenant*
21 that the singer is a tenant. Your credence in *Tenant* is the same as your credence
22 in *Bob*, for given your state of awareness these two propositions are equivalent
23 ... Now let's suppose the possibility suddenly occurs to you that there might be
24 another tenant living in the same flat (*Other*).

25 Initially, you thought the singer could either be the landlord or Bob, the tenant. Then you come
26 to the realization that a third person could be the singer, another tenant. Before awareness
27 growth, that Bob is in the shower and that a tenant is in the shower are equivalent descriptions.
28 After awareness growth, this equivalence breaks down.

⁶Note that $P(H=present|E=seen) \neq P(H=present|R=seen-reported)$, but since you are conditioning on different propositions, this does not conflict with Reverse Bayesianism.

1 Why is this scenario problematic? Suppose, after you hear singing in the shower, you become
2 sure someone is in there, but you cannot tell who. So $P(\text{Landlord}) = P(\text{Bob}) = 1/2$, and since
3 *Bob* and *Tenant* are equivalent, also $P(\text{Tenant}) = 1/2$. Now, *Landlord*, *Bob* and *Tenant* are all
4 propositions that you were originally aware of, and thus Reverse Bayesianism requires that their
5 assigned probabilities should remain in the same proportion after your awareness grows. But
6 note that *Other* entails *Tenant* and *Bob* and *Other* are disjoint, so it follows that $P^+(\text{Other})$
7 must have zero probability.⁷ This is an undesired outcome that rules out the possibility that
8 there could be a third person in the shower.⁸

9 Consider now Mathani's second counterexample:

10 COIN: You know that I am holding a fair ten pence UK coin which I am about to
11 toss. You have a credence of 0.5 that it will land *Heads*, and a credence of 0.5 that
12 it will land *Tails*. You think that the tails side always shows an engraving of a lion.
13 So you also have a credence of 0.5 that (*Lion*) it will land with the lion engraving
14 face-up: relative to your state of awareness *Tails* and *Lion* are equivalent.... Now
15 let's suppose that you somehow become aware that occasionally ten pence coins
16 have an engraving of Stonehenge on the tails side.
17 *Tails* and *Lion* are equivalent propositions prior to awareness growth. Suppose you initially
18 gave *Tails* and *Lion* the same credence. Reverse Bayesianism requires that their relative
19 proportions should stay the same after awareness grow. The same applies to *Heads* and *Tails*.
20 But since *Lion* and *Stonehenge* are incompatible and the latter entails *Tails*, you should have
21 $P^+(\text{Stonehenge}) = 0$, again an undesirable conclusion.

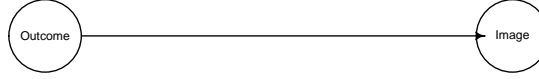
22 Mathani notes that COIN has the same structure as TENANT. This is true to some extent, but
23 there is also an interesting asymmetry between the two scenarios. In TENANT, it is natural to
24 assign 1/3 to *Landlord*, *Bob* and *Other* after awareness growth. That someone is singing in
25 the shower is evidence that someone must be in there, but without any more discriminating
26 evidence, each person should be assigned the same probability. Consequently, a probability
27 of 2/3 should be assigned to *Tenant*. On this picture, the proportion of *Landlord* to *Tenant*
28 changes from 1:1 (before awareness growth) to 1:2 (after awareness growth). But, in COIN,
29 the relative proportion of *Heads* to *Tails* should remain constant throughout, unless evidence

⁷If $P^+(\text{Other}) > 0$, the proportion of *Tenant* to *Landlord* or the proportion of *Bob* to *Landlord* should change.

⁸Awareness Rigidity is no of help either because it would require that $P^+(\text{Landlord}|\text{Landlord} \vee \text{Tenant}) = P^+(\text{Bob}|\text{Landlord} \vee \text{Tenant})$ both equal 1/2, thus forcing $P^+(\text{Other}|\text{Landlord} \vee \text{Tenant})$ to zero.

1 emerges that the coin is not fair. One might have expected that *Landlord* and *Tenant* would
 2 behave just like *Heads* and *Tails*, but actually they do not.

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



6 The upstream node *Outcome* has two states, *tails* and *heads*. These two states remain the same
 7 throughout. What changes are the states associated with the *Image* node downstream. Before
 8 awareness growth, the node *Image* has two states: *lions* and *heads-image*.⁹ You assume that
 9 *Image* = *lions* is true if and only if *Outcome* = *tails* is true. Then, you come to the realization
 10 that the imagines for tails include a lion or a stonehenge engraving. So, after awareness growth,
 11 the node *Image* contains three states: *lion*, *stonehenge* and *heads-image*. Consider now the
 12 other scenario, TENANT. We start with the following graph:



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

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

⁹The heads side must have some image, not specified in the scenario.

1 We conclude with some programmatic remarks. We think that the awareness of agents grows
2 while holding fixed certain material structural assumptions, based on commonsense, semantic
3 stipulations or causal dependency.¹⁰ To model awareness growth, we need a formalism that
4 can express these material structural assumptions. This can be done using Bayesian networks,
5 and we offered some illustrations of this strategy. These material assumptions also guide
6 us in formulating the adequate conservative constraints, and these will inevitably vary on
7 a case-by-case basis. The literature on awareness growth from a Bayesian perspective is
8 primarily concerned with a formal, almost algorithmic solution to the problem. Insofar as
9 Reverse Bayesianism is an expression of this formalistic aspiration, we agree with Steele and
10 Stefánsson that we are better off looking elsewhere.

11 **4 Towards a general theory**

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

20 **5 Counterexamples**

21 In this section, we rehearse two of the counterexamples to Reverse Bayesianism by Steele and
22 Stefánsson. One example targets awareness expansion and the other awareness refinement
23 (more on this distinction soon). We show why they make a limited case against Reverse
24 Bayesianism and then provide a better counterexample with the aid of Bayesian networks.

¹⁰Arrows 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.

P(Image Outcome)		Outcome	
		heads	tails
Image	lion	0	1
	heads-image	1	0
P+(Image Outcome)		Outcome	
		heads	tails
Image	lion	0	1/2
	stonehenge	0	1/2
	heads-image	1	0
P(Outcome) = P+(Outcome)		Outcome	
	heads	tails	
	1/2	1/2	

P(Role Person)		Person		
		landlord-person	bob	
Role	tenant	0	1	
	landlord	1	0	
P+(Role Person)		Person		
		landlord-person	bob	other
Role	tenant	0	1/2	1/2
	landlord	1	0	0
P(Person)		Person		
	landlord-person	bob		
	1/2	1/2		
P+(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.

1 5.1 Friends and Movies

2 The difference between expansion and refinement is intuitively plausible, but can be tricky
3 to pin down formally. A rough characterization will suffice here. Suppose, as is customary,
4 propositions are interpreted as sets of possible worlds, where the set of all possible worlds is
5 the possibility space. An algebra of propositions thus interpreted induces a partition of the
6 possibility space. Refinement occurs when the new proposition added to the algebra induces a
7 more fine-grained partition of the possibility space. Expansion occurs when the new proposition
8 is inconsistent with the existing ones, thus making the old partition no longer exhaustive.

9 The first counterexample by Steele and Stefánsson targets cases of awareness expansion:

10 FRIENDS: Suppose you happen to see your partner enter your best friend's house
11 on an evening when your partner had told you she would have to work late. At
12 that point, you become convinced that your partner and best friend are having
13 an affair, as opposed to their being warm friends or mere acquaintances. You
14 discuss your suspicion with another friend of yours, who points out that perhaps
15 they were meeting to plan a surprise party to celebrate your upcoming birthday—a
16 possibility that you had not even entertained. Becoming aware of this possible
17 explanation for your partner's behaviour makes you doubt that she is having an
18 affair with your friend, relative, for instance, to their being warm friends. (Steele
19 & Stefánsson, 2021, sec. 5, Example 2)

20 Initially, the algebra only contains the hypotheses 'my partner and my best friend met to
21 have an affair' (*Affair*) and 'my partner and my best friend met as friends or acquaintances'
22 (*Friends/acquaintances*). The other proposition in the algebra is the evidence, that is, the fact
23 that your partner and your best friend met one night without telling you (*Secretive*). Given this
24 evidence, *Affair* is more probable than *Friends/acquaintances*:

$$P(Affair|Secretive) > P(Friends/acquaintances|Secretive). \quad (>)$$

25 When the algebra changes, a new hypothesis is added which you had not considered before:
26 your partner and your best friends met to plan a surprise party for your upcoming birthday

1 (*Surprise*). Given the same evidence, *Friends/acquaintances* is now more likely than *Affair*:

$$P^+(Affair|Secretive) < P^+(Friends/acquaintances|Secretive). \quad (<)$$

2 This holds assuming that hypothesis *Surprise* is more likely than the hypothesis *Affair*:

$$P^+(Surprise|Secretive) > P^+(Affair|Secretive),$$

3 and, in addition, that *Surprise* implies *Friends/acquaintances*. After all, in order to prepare a
4 surprise party, your partner and best friend have to be at least acquaintances.

5 The conjunction of ($>$) and ($<$) violates Reverse Bayesianism since *Friends/acquaintances*
6 and *Affair* are basic propositions that do not contain any connectives. But, as Steele and
7 Stefánsson admits, Reverse Bayesianism can still be made to work with a slightly different—
8 though quite similar in spirit—condition, called **Awareness Rigidity**:

$$P^+(A|T^*) = P(A),$$

9 where T^* corresponds to a proposition that picks out, from the vantage point of the new
10 awareness state, the entire possibility space before the episode of awareness growth. In our
11 running example, the proposition $\neg Surprise$ picks out the entire possibility space in just this
12 way. And conditional on $\neg Surprise$, the probability of *Affair* does not change. Thus,

$$P^+(Affair|Secretive \& \neg Surprise) > P^+(Friends/acquaintances|Secretive \& \neg Surprise).$$

13 Awareness Rigidity is satisfied. Reverse Bayesianism—the spirit of it, not the letter—stands.

14 This is not the end of the story, however. Steele and Stefánsson offer another counterexample
15 that also works against Awareness Rigidity, this time targeting a case of refinement:

16 **MOVIES:** Suppose you are deciding whether to see a movie at your local cinema.
17 You know that the movie's predominant language and genre will affect your
18 viewing experience. The possible languages you consider are French and German
19 and the genres you consider are thriller and comedy. But then you realise that,
20 due to your poor French and German skills, your enjoyment of the movie will
21 also depend on the level of difficulty of the language. Since it occurs to you that

1 the owner of the cinema is quite simple-minded, you are, after this realisation,
2 much more confident that the movie will have low-level language than high-level
3 language. Moreover, since you associate low-level language with thrillers, this
4 makes you more confident than you were before that the movie on offer is a thriller
5 as opposed to a comedy. (Steele & Stefánsson, 2021, sec. 5, Example 3)

6 This is a case of refinement. For you initially categorized movies by just language and genre,
7 and then you refined your categorization by adding another variable, level of difficulty. Without
8 considering language difficulty, you assigned the same probability to the hypotheses *Thriller*
9 and *Comedy*. But learning that the owner was simple-minded made you think that the level
10 of linguistic difficulty must be low and the movie most likely a thriller rather than a comedy
11 (perhaps because thrillers are simpler—linguistically—than comedies). So, against Reverse
12 Bayesianism, MOVIES violates the condition $\frac{P(Thriller)}{P(Comedy)} = \frac{P^+(Thriller)}{P^+(Comedy)}$.

13 The counterexample also violates Awareness Rigidity. For consider a proposition that picks
14 out the entire possibility space, for example, $Thriller \vee Comedy$.¹¹ Awareness Rigidity would
15 require that $P(Thriller) = P^+(Thriller|Thriller \vee Comedy)$. But MOVIES does not satisfy this
16 equality since the probability of *Thriller* has gone up.

17 How good of a counterexample is this? Steele and Stefánsson consider an objection:

18 It might be argued that our examples are not illustrative of ... a simple growth in
19 awareness; rather, our examples illustrate and should be expressed formally as
20 complex learning experiences, where first there is a growth in awareness, and then
21 there is a further learning event ... In this way, one could argue that the awareness-
22 growth aspect of the learning event always satisfies Reverse Bayesianism.

23 Admittedly, MOVIES can be split into two episodes. In the first, you entertain a new variable
24 besides language and genre, namely the language difficulty of the movie. In the second episode,
25 you learn something you did not consider before, namely that the owner is simple-minded.
26 Could Reverse Bayesianism still work for the first episode, but not the second? Steele and
27 Stefánsson do not address this question explicitly, but insist that no matter the answer both
28 episodes are instances of awareness growth. We agree with them on this point. Awareness
29 growth is both *entertaining* a new proposition not in the initial awareness state of the agent

¹¹Since MOVIES is a case of refinement, $Thriller \vee Comedy$ picks out the entire possibility space both before and after awareness growth.

1 and *learning* a new proposition. Nonetheless, many could still wonder. Is the second episode
2 (learning something new) necessary for the counterexample to work together with the first
3 episode (mere refinement without learning)?

4 Suppose the counterexample did work only in tandem with an episode of learning something
5 new. If that were so, defenders of Reverse Bayesianism or Awareness Rigidity could still
6 claim that their theory applies to a large class of cases. It applies to cases of awareness
7 refinement without learning and also to cases of awareness expansion. For recall that the
8 first putative counterexample featuring awareness expansion—FRIENDS—did not challenge
9 Reverse Bayesianism insofar as the latter is formulated in terms of its close cousin, Awareness
10 Rigidity. So the force of Steele and Stefánsson’s counterexamples would be rather limited.

11 References

- 12 Bradley, R. (2017). *Decision theory with a human face*. Cambridge University Press.
13 Chihara, C. S. (1987). Some problems for bayesian confirmation theory. *British Journal for*
14 *the Philosophy of Science*, 38(4), 551–560.
15 Earman, J. (1992). *Bayes or bust? A critical examination of bayesian confirmation theory*.
16 MIT press.
17 Glymour, C. (1980). *Theory and evidence*. Princeton University Press.
18 Howson, C. (1976). The development of logical probability. In *Essays in memory of imre*
19 *lakatos. Boston studies in the philosophy of science* (pp. 277–298). Springer.
20 Karni, E., & Vierø, M.-L. (2015). Probabilistic sophistication and reverse bayesianism. *Journal*
21 *of Risk and Uncertainty Volume*, 50, 189–208.
22 Lakatos, I. (1968). Changes in the problem of inductive logic. *Studies in Logic and the*
23 *Foundations of Mathematics*, 51, 315–417.
24 Mathani, A. (2020). Awareness growth and dispositional attitudes. *Synthese*, 198(9), 8981–
25 8997.
26 Pettigrew, R. (forthcoming). How should your beliefs change when your awareness grows?
27 *Episteme*, 1–25. <https://doi.org/10.1017/epi.2022.33>
28 Roussos, J. (2021). Awareness growth and belief revision. *Manuscript*.
29 Schaffer, J. (2016). Grounding in the image of causation. *Philosophical Studies*, 173, 49–100.
30 Steele, K., & Stefánsson, O. (2021). Belief revision for growing awareness. *Mind*, 130(520),
31 1207–1232.
32 Wenmackers, S., & Romeijn, J.-W. (2016). New theory about old evidence: A framework for
33 open-minded bayesianism. *Synthese Volume*, 193, 1225–1250.
34 Williamson, J. (2003). Bayesianism and language change. *Journal of Logic, Language, and*
35 *Information*, 12(1), 53–97.
36 Zabell, S. (1992). Predicting the unpredictable. *Synthese*, 90(1), 205–232.