Meaning, shifts and conditioning
Romeijn, J.-W.

Published in:
Default journal

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2009

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Meaning shifts and Conditioning

Jan-Willem Romeijn
Department of Philosophy
University of Groningen
j.w.romeijn@rug.nl

Abstract
This paper investigates the viability of the Bayesian model of belief change. Van Benthem (2003) has shown that a particular kind of information change cannot be modelled by Bayesian conditioning. I argue that the problems come about because the information change alters the semantics in which the change is supposed to be modelled by conditioning: it induces a shift in meanings. I then show that meaning shifts can be modelled as conditioning by employing a semantics that allows to make these changes in meaning explicit, and that the appropriate probability kinematics can be described by Dempster’s rule.

1 Conditioning and meaning shifts

In the Bayesian model of belief change, beliefs over sentences, like $s$ and $r$, are represented with probability assignments over an algebra of propositions, $S$ and $R$, which are associated with the sentences by an interpretation: a propositions fixes the meaning of a sentence. A belief change due to the acceptance of a sentence is reflected in the probability assignment by conditioning the assignment on the proposition associated with the sentence, as determined by Bayes’ rule,

$$p_S(R) = p(R|S), \quad (1)$$

meaning that the new probability assignment to $R$ after learning $S$ is the old assignment to $R$ conditional on $S$. The probability measures over sets of possible worlds $R$ and $S$ may be read as degrees of belief in the sentences $r$ and $s$. So we can also write $p_s(r) = p(r|s)$; upper and lower case can be used interchangeably.
The idea of conditioning is that within the set of possible worlds consistent with the sentence that is learnt, the probability is kept unchanged. This conservativity of belief change is usually referred to as rigidity. It seems rather natural to impose this condition if we interpret sentences as propositions: after learning the sentence $s$, we do not consider possible those worlds in which $\neg s$ holds, i.e., we zoom in on the worlds within the set $S$. But apart from that there is no reason to favour one world in which $s$ holds over another such world. These intuitions can also be given a further underpinning in the so-called dynamic Dutch book arguments. See Jeffrey (1992).

Since we use propositions, or sets of possible worlds, to fix the meaning of sentences, a shift in the meaning of a sentence can be represented by a change in the associated proposition, that is, in the collection of possible worlds that is associated with the sentence. Now imagine that the meaning of a sentence is partly determined by the set of sentences that have been accepted thus far. Then it may happen that some sentence $s$ is accepted, but that at the same time the acceptance of $s$ causes the sentence $r$ to shift in meaning. In other words, before the acceptance of $s$ the sentence $r$ is associated with one set of possible worlds, $R$, but after the acceptance of $s$ it is associated with a different set of possible worlds, $R'$ say. As will be illustrated below, such meaning shifts occur in contexts where people reason about knowledge, and perform so-called epistemic actions. This is the domain of epistemic logic.

Shifts in meaning of the above kind cause problems for the Bayesian model. Imagine that all possible worlds associated with the sentence $s$ are worlds that satisfy the sentence $r$, so that accepting $s$ entails that we assign $r$ a probability 1. However, accepting $s$ may change the meaning of $r$, so that after the meaning shift some worlds that previously satisfied both $s$ and $r$, and that had nonzero probability before the update, fail to satisfy $r$ after all. In that case the probability for $r$ must be smaller than 1 after a complete update with $s$. In other words, meaning shifts lead to a conflict between conditional and updated probability.

2 An example from epistemic logic

Van Benthem (2003) has shown that some cases of belief change cannot be modelled by Bayesian conditioning. In this section I reiterate his example
Imagine Alice and Bob, who are both investigating the state of the world they are inhabiting. The Kripke semantics of Figure 1 summarises the possible worlds and which of these they can tell apart. There are three dots which represent worlds: world 1 in which $s \land \neg r$, world 2 in which $s \land r$, and world 3 in which $\neg s \land r$. The lines indicate so-called epistemic accessibility. If world 1 is actual, Bob knows that $s \land \neg r$: he only considers it possible that he is in world 1. Meanwhile, in world 1, Alice knows that $s$ but she does not know whether $r$, because she also considers it possible that she is in world 2, and while $\neg r$ holds in world 1, $r$ holds in world 2. This also means that in world 1 Alice does not know that Bob knows that $s \land r$, because in that case she would know as well. In world 2, the situation looks much the same for Alice. Bob, on the other hand, knows that $r$ but he does not know whether $s$, because he considers both world 2 and world 3 possible. Again Alice and Bob do not know of each other’s ignorance, which would allow them to conclude they inhabit world 2. In world 3, finally, Alice knows that $\neg s \land r$ since she only considers world 3 possible, while for Bob everything looks much the same as in world 2.

Note that sentences like $r$ and $s$ are all associated with possible worlds, given by the dots, and not with the epistemic relations as summarised by the double arrows. Sentences about knowledge, on the other hand, are concerned also with the arrows. Note further that as it is described, Alice and Bob have a rather shallow epistemic relation. They take notice of their own beliefs about the worlds and of each other’s beliefs about the world, but they do not take notice of each other’s beliefs about each other’s beliefs, or
of any knowledge of higher order than that. On the other hand, the Kripke diagram does summarise the higher-order beliefs that Alice and Bob have, concerning each other’s beliefs and so on, because epistemic accessibility relations can be concatenated.

Finally, note that because Alice and Bob both have the knowledge summarised in the above diagram, they can in some worlds discover where they are by mere introspection. Some soul-searching may tell Alice that actually she knows that \( \neg s \), and with the diagram she may then conclude that she is in world 3. In the present setup, such introspections are just another source of information. In modeling the beliefs of the agents, we may suppose information obtained by introspection comes only after a deliberate observation of one’s own state of mind.

Now consider the following sentences: "Alice does not know whether \( r \)”, written as \( \neg K_A r \), and "Bob does not know whether \( s \)”, denoted \( \neg K_B s \). As depicted in Figure 2, the meaning of the sentence \( \neg K_A r \) is the proposition or, equivalently, the set of worlds \( \{ W_1, W_2 \} \), while \( \neg K_B s \) is associated with the proposition \( \{ W_2, W_3 \} \). Now say that we assign all worlds a prior probability \( p(W_i) = \frac{1}{3} \), and imagine that we publicly announce to both Alice and Bob that Alice does not know \( r \), \( \neg K_A r \). With this information Alice and Bob both turn to the diagram, and conclude that they cannot be in \( W_3 \), because in that world Alice knows that \( r \). According to a naive Bayesian model, and as suggested by Figure 3, the new probability assignment \( p_{\neg K_A r} \) for the proposition \( \neg K_B s \) must therefore be \( p_{\neg K_A r}(\neg K_B s) = p(\neg K_B s|\neg K_A r) = p(\{ W_2, W_3 \}|\{ W_1, W_2 \}) = \frac{1}{2} \).

But this is not the whole story about the update. Apart from ruling out \( W_3 \), a complete update requires that Alice and Bob also operate on the
epistemic relations between the remaining worlds. Specifically, because $W_3$ is ruled out, neither of them will include epistemic relations with $W_3$ after the update. So from world 2 there is no accessibility relation that Bob has with world 3 anymore. The proper representation of the new situation is therefore given in Figure 4. Importantly, in the new epistemic situation Bob is not in doubt about the sentence $s$ in any of the two remaining worlds. After the complete update, there are no possible worlds associated with the sentence $\neg K_B s$. The new probability must therefore be $p_{\neg K_A r}(\neg K_B s) = p(\emptyset|\{W_2, W_3\}) = 0$, and not the half derived earlier.

This shows that simply conditioning on the proposition $\{W_1, W_2\}$ leads to the wrong probability assignment for the sentence $\neg K_B s$. In terms of the preceding section, and as suggested in Figure 4, we might say that due to the acceptance of $\neg K_A r$ the meaning of $\neg K_B s$ shifts: the associated proposition changes from $\{W_2, W_3\}$ to $\emptyset$. Learning the sentence $\neg K_A r$ not only restricts the set of worlds that Alice and Bob consider possible, but also operates
on the epistemic relations between worlds, to the effect that the remaining worlds themselves are altered.

3 The epistemic update as conditioning

It seems that the Bayesian model of belief change simply cannot accommodate the above case of learning: meaning shifts may cause a conflict between conditional and updated probability, and thus cause violations of rigidity. One may therefore conclude, with van Benthem, that Bayesian conditioning is a model with limitations, and subsequently turn to an entirely different formal model of belief change. The reaction of this paper is somewhat more positive. It is to look for a characterisation of possible worlds and a generalisation of Bayesian updating that together allow us to incorporate the notion of a meaning shift.

The guiding idea is that all the aspects of the information that play a part in the event of learning new information must somehow be made explicit in the semantics. The foregoing employed a thin notion of possible world, characterised only by the propositions $r$ and $s$ being true or false in them, to which the epistemic structure may be added by means of accessibility relations between worlds. The following, by contrast, assumes a thick notion of possible worlds, which are also characterised by the epistemic properties given by propositions such as $\neg K_B s$. As I will illustrate, this idea of thick worlds is crucial to a more Bayesian treatment of belief change. This view is inspired by Halpern (2004), who also maintains that all the aspects of the protocol that is followed in providing an agent with information must somehow find their way into the possible worlds semantics that is used for framing the information changes.

We need not unfold the Kripke structure of Figure 1 very far to arrive at a semantics that can accommodate the learning events described in the foregoing. Instead of distinguishing only the possible worlds 1, 2 and 3, we now make a further distinction between the possible epistemic situations that Alice and Bob can conceive of within each of the possible worlds. Within world 1, for instance, Bob can only conceive of world 1, while Alice considers world 1 and world 2 possible. Accordingly, we distinguish two epistemic situations that both belong to world 1: in both of them Bob thinks he is in world 1, while in one of them Alice thinks she is in world 1, and in the
other she thinks she is in world 2. Neither of them can distinguish these situations. Similarly, we can distinguish four epistemic situations in world 2: Alice thinks she is in world 1, or 2, and Bob thinks he is in world 2, or 3. World 3 has two epistemic situations again, corresponding to Bob thinking he is in world 2, or 3. In analogy to Figure 1, Figure 5 summarises the new set of possible worlds and situations.

It is now easy to model the inclusion of the information expressed by the sentence that Alice is in doubt over \( r \), \( \neg K_A r \), by means of conditioning. We say that world \( i \) is one in which Alice does not know \( r \) if within this world we find epistemic situations that vary with respect to whether Alice holds \( r \) true or not. Accordingly, an epistemic situation in a world is associated with Alice not knowing \( r \) if there are, within the same world, other situations that are different for Alice with respect to \( r \). Conditioning on the sentence \( \neg K_A r \) means that we zoom in on those worlds at which Alice does not know \( r \). In this case, as illustrated by Figure 6, we eliminate world 3 from the possible world semantics. This first stage of the update corresponds to the simple Bayesian update discussed in the foregoing: after the update we have only world 2 left in which Bob does not know \( s \), so that presumably \( p_{\neg K_A r}(\neg K_B s) = 1/2 \). This update is equivalent to the update depicted in Figure 3.

In the foregoing it was shown that the update operation of eliminating world 3 is not the full story, because this elimination interferes with the epistemic accessibility relations, leading to a meaning shift. In the present setup this additional operation can also be modelled as conditioning. Because we have eliminated world 3, we must also remove all epistemic situations from
Figure 6: The possible worlds semantics depicting the epistemic situations for Alice and Bob after updating with the information that they are not in world 3, but without Alice of Bob having thought this information through.

Figure 7: The possible worlds semantics depicting the epistemic situations for Alice and Bob after updating with the full information included in $\neg K_A r$.

other worlds in which Alice or Bob think they are in world 3. That is, we do not just condition on the information that Alice and Bob are not in world 3, we condition on the full information that Alice does not know $r$, namely that world 3 is excluded, but thereby also that Alice and Bob cannot think they are in world 3. In this second stage of the update, Alice and Bob work out the epistemic consequences of what they learnt in the first stage.

The update that corresponds to this stage is depicted in Figure 7, which is analogous to the update of Figure 4. Note, however, that in the semantics of Figure 5, this update is a conditioning operation: we eliminate the two epistemic states in world 2 in which Bob thinks he is in world 3. In this possible worlds semantics, there are no worlds left in which Bob does not know $s$, so that we indeed have that $p_{\neg K_A r}(\neg K_B s) = 0$. 
4 Epistemic situations as possible worlds?

In the above update we have conditioned on a particular set of epistemic situations, namely the complement of the union of three sets of epistemic situations: all situations within world 3, all situations in which Alice thinks she is in world 3, all situations in which Bob thinks he is in world 3. In the cube of Figure 5, we have removed the layer labeled 3 in all three dimensions. We can say that the set of situations conditioned upon is the information contained in the sentence $\neg K_A r$. But did we thereby avoid the problem with Bayesian conditioning of Section 2? Answering this question requires us to carefully separate the notions of proposition, information, and meaning.

It may seem that we can resolve the paradox simply by a semantics of epistemic situations. We may say that the epistemic situations are the units in which we interpret the sentences, i.e., give them propositional content. Accordingly, we divide the probability equally over these situations: each situation receives a probability of $1/8$. In addition, we stipulate that the proposition associated with the sentence $\neg K_A r$ is the set of epistemic situations within worlds 1 and 2, and nothing more, and further that the proposition associated with the sentence $\neg K_A r$ is the set of epistemic situations within worlds 2 and 3. However, in that case we have that $p_{\neg K_A r}(\neg K_B s) = 2/3$. This would be in accordance with what Alice and Bob supposedly learnt because, by the explicit meaning of $\neg K_A r$ in terms of situations, they learnt that they are not in world 3, but they did not reflect any further on that. On this view, then, the introspective part to the belief change is simply missing, and the information change that goes with it must be modelled as a separate update.

Furthermore, when it comes to that introspective part, a Bayesian model in which the epistemic situations are taken as basic comes out wrong. To see this, say that the proposition associated with the sentence $\neg K_A r$ is the complement of the set of epistemic situations labeled 3 in all three dimensions, as proposed in Section 3. This meaning of $\neg K_A r$ expresses that Alice and Bob think through the implications of the sentence $\neg K_A r$ once it is learnt. The sentence $\neg K_B s$, on the other hand, may be associated with the set of all epistemic situations in worlds 2 and 3, because in these worlds Bob does not know $s$. But then, after a Bayesian update on $\neg K_A r$, we find that $p_{\neg K_B r}(\neg K_A s) = 1/2$, which is what we were trying to avoid. Alternatively,
we might say that the sentence $\neg K_B s$ rules out the situations within world 1, but also all other situations in which world 1 features, in analogy to the meaning of $\neg K_A r$. In that case, we find that $p_{\neg K_B r}(\neg K_A s) = 1/4$ after the update!

These answers are clearly not satisfactory as part of the intended model of belief change. In fact, it seems that we arrive back at square one with regard to modelling meaning shifts. In the Kripke semantics of Figure 1, whether a world is included in a certain proposition depends on the relations this world has with other worlds: it is by virtue of the epistemic accessibility relation between worlds 2 and 3 that we say that world 2 belongs to $\neg K_B s$. The whole problem with the epistemic example can traced back to such dependencies of worlds on each other. But the same interdependency holds for the epistemic situations: the situation within world 2 in which Bob thinks he is in world 2 is one in which he is in doubt by virtue of the fact that, within world 2, there is also a situation in which he thinks he is in world 3. By simply taking epistemic situations as the units of analysis, we do not rid ourselves of this aspect of epistemic updates, and more generally, we do not manage to capture the meaning shifts in a formal model of belief change.

5 Propositions, information, and meaning shifts

In the following, we will therefore not take propositions as sets of epistemic situations. Rather we take propositions as sets of worlds again, so that meaning pertains to sets of worlds, while each world in turn comprises of a set of, what will be called, possibilities within worlds. In the epistemic example, these possibilities are epistemic situations. The worlds in the example are thus characterised by, first of all, truth valuations over basic sentences such as $r$ and $s$ and, secondly, by the truth valuations of sentences involving these possibilities, namely the epistemic situations. For example, the sentence $r$ is associated with worlds 1 and 2, because this sentence is true in those worlds. And the sentence $\neg K_B s$ is associated with worlds 2 and 3, because within these worlds we find epistemic situations over which the valuation that Bob assigns to $s$ vary, and hence $\neg K_B s$ is true in these two worlds. This understanding of worlds is effectively identical to how worlds are conceived of in a standard Kripke semantics.
Now on this view, we cannot maintain Bayes’ rule, Equation (1), as the rule that governs all belief change. After all, we then have that \( p(\neg K_B s | \neg K_A r) = p(W_2, W_3 | W_1, W_2) = 1/2 \) but we want \( p_{\neg K_A r}(\neg K_B s) = 0 \). Apparently, in the transition from \( p(\cdot) \) to \( p_{\neg K_A r}(\cdot) \) we make use of more than just the propositional content of \( \neg K_A r \): getting from \( p(\cdot) \) to \( p(\cdot | \neg K_A r) \) is not enough.

In the following I elaborate on how the framework presented in Section 3 can be used to make this additional content precise, and how it allows us to view the learning involved in it as conditioning, albeit not Bayesian conditioning.

It will be useful to get clear on the distinction between propositional content and information, which we will here discuss in terms of the epistemic example. The propositional content of the sentence \( \neg K_A r \) is given by the worlds \{W_1, W_2\}. This content is also called the meaning of the sentence. The information contained in the sentence \( \neg K_A r \), on the other hand, can be spelled out in terms of epistemic situations. It is conceivable that both Alice and Bob do not think the sentence \( \neg K_A r \) through, in which case it presents them with the information of all epistemic situations within worlds 1 and 2, eliminating only the situations in world 3. But after some introspection, Alice and Bob may realise that they cannot conceive of themselves in world 3 anymore, so that Bob can eliminate the two epistemic situations in world 2 in which he thinks he is in world 3. The information expressed by \( \neg K_A s \) then consists in the complement of all the epistemic situations labeled 3 in one or more dimensions.

With the distinction between propositional and informational content in place, we can make precise what is learnt when accepting the sentence \( \neg K_A s \). In moving from the probability assignment \( p(\cdot) \) to the assignment \( p_{\neg K_A r}(\cdot) \), we do not just condition on the propositional content, \( p(\cdot | \neg K_A r) \), but rather we condition on the information contained in the sentence \( \neg K_A r \). This conditioning operation models a genuine meaning shift: the information in the sentence \( \neg K_A s \) induces a change to world 2, so that the sentence \( \neg K_B s \), whose propositional content included world 2 before the information \( \neg K_A r \) came in, undergoes a change, since it does not include world 2 after the information \( \neg K_A r \) has been processed. It is the conditioning on epistemic situations within worlds that causes this shift in propositional content, or in other words, a shift in meaning.

This brings out the advantage of the representation of Figure 5, in which propositions and information come apart. We express beliefs by a proba-
bility assignment over sets of worlds, or propositions, but we frame the information we receive in terms of sets of epistemic situations, which may cut across worlds. Information change is modelled by means of conditioning on information, but this may involve operating within the worlds. As a result, conditioning on information may require us to redraw the map of propositions. Conditioning thus captures the way in which the meanings of sentences are changed.

6 Probability kinematics for meaning shifts

In the foregoing we redescribed the meaning shift of the example in terms of conditioning on sets of epistemic situations. It was seen that this models meaning shifts because such sets can cut across possible worlds. But the Bayesian model concerns the probability assignments over propositions, and hence over the worlds alone. Therefore the Bayesian model cannot specify how these probability assignments react to the meaning shifts. To define the probability kinematics for meaning shifts, we introduce some new formal machinery.

We first generalise the semantics of epistemic situations defined in the foregoing. The idea is that possible worlds have an internal structure, and that they therefore consist in a set of elementary possibilities. In the epistemic example, the elementary possibilities concern the beliefs of the agents. But other examples of meaning shifts may require a different notion of elementary possibility. These possibilities are denoted with $w_i$, worlds are sets of possibilities $W_k$. It is crucial that we may learn information that is expressed by sets of possibilities that cut across worlds, i.e., sets $V$ that cannot be spelled out as a conjunction of worlds $W_k$.

We now model the epistemic update by weakening probability assignments to Dempster-Shafer belief functions over these elementary possibilities. As will be seen, this allows for a semi-Bayesian model of the belief change. A more extensive discussion of Dempster-Shafer belief functions can be found in Halpern (2003, pp. 32–40). Belief functions are based on a measure over the space of possible possibilities, a mass function $m$, that is only defined for the subalgebra of worlds, namely $\mathcal{W} = \mathcal{P}(W)$ with $W = \{W_k : k = 1, 2, \ldots n\}$. We have $m(U) \in [0, 1]$ for all members of $U \in \mathcal{W}$, and $\sum_k m(W_k) = 1$, but for any other set $V \notin \mathcal{W}$ the function
\( m(V) \) is not defined. From the mass function we can derive an imprecise probability assignment for any \( V \) according to the formula:

\[
\max_{U \subseteq V} m(U) \leq p(V) \leq \min_{U' \subseteq V'} m(U'),
\]

(2)

for \( U \in \mathcal{W} \). So any set of situations \( V \) receives a probability that is larger than the largest mass defined for a set \( U \) included in \( V \), and smaller than the smallest mass defined for any of the sets \( U' \) in which \( V \) is included.

But how do we incorporate new information into this imprecise or partial probability assignment? The answer is that we determine a new mass function by means of Dempster’s rule of combination. We first define another mass function \( m^* \) that expresses the information. In principle this can be any mass function, assigning masses to elements from a partition \( \mathcal{V} \). In the special case in which the information is presented as a particular set in the space, \( V \), then the mass function is simply defined as \( m^*(V) = 1 \) and \( m^*(\overline{V}) = 0 \), where \( \overline{V} \) is the complement of \( V \). In general, the mass function after the update is given by

\[
m_V(W_k \cap V_{k'}) = (m \oplus m^*)(W_k \cap V_{k'}) = \frac{m(W_k)m^*(V_{k'})}{\sum_{k''}m(W_km^*(V_{k''}))}.
\]

If \( m^* \) is a function putting all mass on a single set \( V \), we write \( m_V \) for the new mass function. An update to \( m_V \) is fairly straightforward: for any \( k \) such that \( W_k \cap V \neq \emptyset \), we have \( m_V(W_k \cap V) \propto m_{W_k} \), while \( m_V(W_k \cap \overline{V}) = 0 \).

### 7 Properties of belief functions

Let me emphasise a number of properties of the belief functions. First of all, the probabilities of elementary possibilities within the worlds \( W_k \) are not given, meaning that any probability assignment of the epistemic situations within some world that sums to the probability assigned to the world as a whole is allowed. Every mass function thus corresponds to a set of probability assignments \( \mathcal{P} \) that comply to the restrictions set by Equation (2). These restrictions are that

\[
p(\{w : w \in W_k\}) = m(W_k)
\]

for all \( p \in \mathcal{P} \). For more on the precise formal relation between belief functions, inner and outer measures, and sets of probability assignments, I refer to Halpern (2003) and the references in there.
Secondly, we may note that Dempster’s rule of combination is in some sense a refinement of Bayesian updating. To see the connection, we represent a mass function by a set of probability assignments \( P \) over all elementary possibilities. Now imagine we learn the information corresponding to the set of possibilities \( V \), and consider the operation we must perform on the set \( P \) such that the set of updated probability assignments, \( p_V \in P_V \), is again a representation of the mass function \( m_V \). Now if the new information is expressed in a single set \( V \in W \), then the new mass function derived by Dempster’s rule will correspond to the set of probability assignments we arrive at performing a Bayesian update with the set \( V \) on each of the members of \( P \). For all those \( W_k \not\in V \) we have \( m_V(W_k) = 0 = p(W_k|V) \), while for \( W_k \in V \), the mass will be \( m_V(W_k') = 1 \times p(W_k')/p(V) = p(W_k'|V) \) as well. Since there are no operations within worlds, the update is effectively identical to the update we would have had if the worlds \( W_k \) had been the elementary possibilities.

Furthermore, if the set \( V \) cuts across worlds, then for all those \( U \in W \) that do not depend on the internal structure of worlds in terms of elementary possibilities, and hence are not susceptible to meaning shifts, the new probability corresponds with what Bayesian updating prescribes as well. To see this, simply consider the mass and probability assignment on the level of worlds again. For any world \( W_k \) not intersecting with \( V \), we have that \( p_V(W_k) = 0 = p(W_k|V) \). For worlds \( W_k' \) intersecting with \( V \), we have that

\[
m_V(W_k') = \frac{1 \times p(W_k')}{p(V')} = p(W_k'|V'),
\]

where \( V' = \cup_{k'} W_{k'} \), the union of all worlds intersecting with \( V \). The new mass function \( m_V \) thus retains the proportions among the masses \( m(W_k') \) for all worlds \( W_k' \) intersecting with \( V \), thereby satisfying the rigidity condition inherent in Bayes’ rule.

Thirdly, and related to the former, there is a connection between the present application of Dempster’s rule, and a particular application of Lewis’ idea of updating by imaging. If the set \( V \) is not a proposition, \( V \not\in W \), learning the information of \( V \) means that the probability of all \( w \not\in V \) is set to 0 by Dempster’s rule. On the other hand, for all \( W_k \) for which \( W_k \cap V \neq \emptyset \), we must have that \( p_V(\{w : w \in W_k\}) = m_V(W_k) \). So the probability of any elementary possibility \( w \not\in V \) which belongs to one of the \( W_k \) intersecting with \( V \) must be projected onto the \( w' \in W_k \) that have \( w' \in \)
In other words, updating by Dempster’s rule in this context is the same as updating by imaging under the condition that elementary possibilities that receive probability 0 divide their probability equally over the remaining possibilities in their world, and that worlds that receive probability 0 divide their probability equally over all remaining worlds.

Finally, we want to remark on those sentences \( u \) whose extension \( U \in \mathcal{W} \) depends on the internal structure of the worlds, or in other words, on the elementary possibilities included in it. In the epistemic example of this paper, \( \neg K_A r \) is such a sentence. As expected, we cannot say that the probability of the corresponding proposition is governed by Bayes’ rule or by Dempster’s rule. This is impossible because of the explicit change in the meaning of \( u \). Still, apart from the change in meaning that is governed by the conditioning on the information contained in \( u \), the probability assignment \( p_U \) is determined entirely by Dempster’s rule. In particular, it is not a problem that we update on a set of possibilities that is not a proposition, and whose probability is not defined.

### 8 Application to the epistemic example

By way of illustration, we now apply the machinery of belief functions to the problem from epistemic logic. But I first introduce some notation.

In the example, the space of elementary possibilities is the set of epistemic situations, which can be partitioned according to worlds, and according to epistemic states. We call the partition into worlds \( \mathcal{W} = \{W_1, W_2, W_3\} \), and the partition into epistemic states \( \mathcal{B} = B_{ij} : i, j = 1, 2, 3 \), in which \( B_{ij} \) signifies that Alice believes she is in world \( i \) while Bob believes he is in world \( j \). The entire space is given by \( \mathcal{W} \times \mathcal{B} \). The information contained in some sentence \( u \) is always a subset of situations \( \langle W_k, B_{ij} \rangle \). For example, the information in the sentence \( K_A r \) is the set \( \{\langle W_3, B_{32} \rangle, \langle W_3, B_{33} \rangle\} \).

In principle any set of epistemic situations counts as information that we may learn. But for the epistemic example, we can make the standard package of information a bit more precise. We will use the convention that \( W_u \) is the set of all worlds \( W_k \) for which \( u \) is true, so for example \( W_{\neg K_A r} = \{W_1, W_2\} \). Note that this replaces the earlier convention, for which the worlds associated with \( u \) were denoted \( U \) and so on. We also write \( B_u \) for the set of all epistemic states that do not involve worlds excluded by \( W_u \). That is, \( B_u \) consists of
all \( B_{ij} \) such that, for all \( k \) for which \( W_k \not\in W : u \), we have \( i \neq k \neq j \). So for example, \( B_{\neg K_A r} = \{ B_{ij} : i, j = 1, 2 \} \) because \( W_{\neg K_A r} \) excludes \( W_3 \), and hence the epistemic states \( B_{33} \) and \( B_{3j} \) are also excluded. As a result, the expression \( W_u \times B_u \), or \( WB_u \) for short, is the set of all epistemic situations that are left after having conditioned on the worlds consistent with \( u \), namely \( W_u \), and after having processed the epistemic repercussions of that, namely by eliminating the states involving worlds outside \( W_u \) or, equivalently, by conditioning on \( B_u \). The set \( WB_{\neg K_A s} \), for example, is the smaller cube on the right of Figure 7.

With this notation in place we can make precise how the Dempster-Shafer belief functions can be applied. We assign a mass function \( m \) to the following events in \( W \times B \):

\[
\begin{align*}
m(W_1) &= m(\{(W_1, B_{11}), (W_1, B_{21})\}) = \frac{1}{3} \\
m(W_2) &= m(\{(W_2, B_{21}), (W_2, B_{31}), (W_2, B_{22}), (W_2, B_{32})\}) = \frac{1}{3} \\
m(W_3) &= m(\{(W_3, B_{23}), (W_3, B_{33})\}) = \frac{1}{3}
\end{align*}
\]

Following Equation 2, the probability of elements like \( (W_1, B_{23}) \) is thus set to zero. But we also have

\[
0 \leq p(\{(W_1, B_{21})\}) \leq \frac{1}{3},
\]

\[
\frac{1}{3} \leq p(\{(W_1, B_{11}), (W_1, B_{21}), (W_2, B_{22})\}) \leq \frac{2}{3}.
\]

As indicated before, the probabilities of the situations within the worlds are not given, meaning that any probability assignment of the epistemic situations within some world that sums to the probability assigned to the world as a whole is allowed.

Now say that we learn the information of the sentence \( \neg K_A r \), and thus condition on the set \( W : \neg K_A r \times B : \neg K_A r \). How do we incorporate this into the mass function? First we determine a mass function \( m^* \) that expresses the information in \( \neg K_A r \): it simply assigns mass 1 to the set \( W : \neg K_A r \times B : \neg K_A r \). The mass function after the update is then defined by Dempster’s rule. After the update, we have

\[
\begin{align*}
m_{\neg K_A r}(-r) &= m_{\neg K_A r}(\{(W_1, B_{11}), (W_1, B_{21})\}) = \frac{1}{2} \\
m_{\neg K_A r}(-r) &= m_{\neg K_A r}(\{(W_2, B_{22}), (W_2, B_{12})\}) = \frac{1}{2}.
\end{align*}
\]

The resulting probability assignment has \( p_{\neg K_A r}(-r) = p_{\neg K_A r}(r) = \frac{1}{2} \), as expected. On the level of worlds, where the mass function is identical to
the probability assignment, Dempster’s rule gives exactly the same results as Bayesian conditioning.

The difference is of course that Dempster’s rule allows us to condition on sets of situations that cut across worlds, whereas Bayesian conditioning is only defined when we condition on sets that correspond to propositions. This is how we can also model the changes in meaning. In the case at hand, we can compute the probability of the sentence \( \neg K_{Bs} \), interpreted as the proposition \( W: \neg K_{Bs} = \{W_2, W_3\} \), conditional on the information contained in the sentence \( \neg K_{Ar} \), namely the set \( W: \neg K_{Ar} \times B: \neg K_{Ar} \). After this, the sentence \( \neg K_{Bs} \) must interpreted as the proposition \( W: \neg K_{Bs} = \emptyset \).

Similarly, we may ask for the probability assignment over the information \( W: \neg K_{Bs} \times B: \neg K_{Bs} \) before and after the conditioning, for which Dempster’s rule determines an interval probability.

Summing up, the formal model of meaning shifts consists of two parts: particular rules on set membership determine the propositional and informational content of sentences, and Dempster’s rule determines how to assign precise and imprecise probabilities to all these sets after conditioning. The latter rule is versatile enough to accommodate the type of sets that enable us to formalise meaning shifts with the former.

9 Further research

With the illustration I hope to have given an idea of the applications that the formal model allows for. The crucial aspect of the model is that it makes a distinction between propositional and informational content. The latter hinges on what I have in the foregoing called a thick notion of possible worlds. In the example, the epistemic facts are not expressed in a structure of accessibility relations between worlds, but rather in terms of epistemic situations within the worlds. This is what allows us to model the meaning shift as conditioning.

Clearly, the model presented in this paper is not the full story on epistemic updates, or on meaning shifts. One application of the present model in the epistemic setting concerns knowledge of higher order than those involved in the example. In principle we can refine the set of epistemic situations within each world further, to include beliefs of Alice about Bob’s beliefs, and so on. The resulting semantics will be much like the so-called
Harsanyi type space that is used in game theory, which allows for a fully Bayesian treatment of higher order beliefs. Brandenburger (2007) provides a very readable introduction to this concept. However, the exact relation between type spaces and the thick worlds employed in this paper must be the topic of another paper.

Another possible extension of the present model employs the fact that Dempster’s rule allows for updates with non-trivial mass functions $m^*$. Instead of an update that mimics Bayesian updating based on the information in a single set $V$, this creates an update similar to an update governed by Jeffrey’s rule, based on a probability assignment over a partition $V$. Effectively we thus redefine the probability over possible information sets. A curious consequence is that information sets whose probability was imprecise may, after the update, receive a precise probability. But how to make sense of these new probabilities is again the subject of another paper.

Finally, a word of caution. The foregoing might suggest that there cannot be a fully Bayesian model for belief changes involving meaning shifts. But this is certainly not true. One option, not investigated in this paper, is to take as possible worlds all evolutions of the worlds and the epistemic situations therein, as for instance given in Figure 5. If we encode the epistemic states from $B$ in a $3 \times 3$-matrix of binary numbers, taking the bits to refer to the different positions $B_{ij}$ within each world $W_k$, a possible world may defined by an entry determining the actual world, namely 1, 2, or 3, together with an infinite sequence of such matrices. In terms of these rather elaborate possible worlds, we can again define an algebra, over which we define probability assignments and operations such as conditioning. This semantics will surely be rich enough for accommodating any belief change over epistemic situations and worlds, because it makes the time evolution of the beliefs explicit in the algebra. However, to be frank, working out the details of such a model might never be the topic of any paper.

References


