The space of possibilities and interactive beliefs

Catrin Campbell-Moore

July 23, 2019

1 Introduction

Someone might believe that the sun will rise tomorrow, or be confident that their train will be on time. In formal epistemology we want to represent this in a formal structure. We take it that the 'agents' have an attitude towards a proposition. A common view is that a proposition is a set of possibilities. This picture underlies most formal work in epistemology. Whilst Frege's puzzle and related challenges has been invoked to argue that there has to be more to the propositions, this underlying picture of propositions as sets of possibilities is pervasive in formal epistemology. Much of this might be because of the interest in credences or degrees of belief, where one uses probability theory. And probability theory starts with an Ω—a space of possibilities; and has a probability function, p, which is defined on subsets of Ω—what we might call 'propositions'. This paper adopts such a perspective, taking the propositions which are the objects of agents beliefs to be collections of possibilities. But we will see that this leads to some specific challenges. If we instead just think of the objects of agents' attitudes as being sentences most of the worries in this paper are circumvented.

This paper questions how this framework looks once we have agents thinking about their own and other agents' beliefs? Formal epistemologists are very interested in interactive beliefs, for example they think about the epistemology of disagreement. We want to allow me to have beliefs about your beliefs, and you to have beliefs about my beliefs. But now does this fit our initial picture? Our initial picture was: there's a space of possibilities, and our beliefs are attached to propositions, understood as sets of the possibilities. How does this fit when the propositions also involve belief facts? We can actually keep this initial picture fixed and fit interactive beliefs into it. The resulting structures are ones that are familiar in (probabilistic) modal logic.

We also investigate a further issue: philosophers often want to simply fix a collection of all the possibilities. They then just think about norms on agents opinions, e.g. saying they should satisfy the axioms of probability. Whilst there are some challenges to this picture (e.g. Hajek, ms) it is a widely held one. And we will adopt this picture of there being a fixed "universal" space of all possible worlds, at least before higher order and interactive beliefs are considered. But we investigate what happens when interactive beliefs are taken into account. This leads to challenges. The challenge arises because what attitudes are possible depends on which worlds are possible, but this depends on which attitudes are possible. We investigate whether this circularity can be avoided and a full specification of all the possibilities and all beliefs over those possibilities can be given.

The answer will ultimately depend on what one thinks it takes to uniquely identify beliefs, and thus what it takes to capture all beliefs. If we impose a relatively strong assumption on the identification of beliefs: that all that matters is their attitudes towards propositions that are expressible in a finitary way, then we will be able to find a universal space that contains all possible beliefs. But without such a strong assumption on the identity of beliefs, if we think attitudes towards all propositions matter, then there will be no universal space. By making this move, though, it looks like we essentially gave up on the idea that the objects of agents attitudes are propositions conceived of as sets of possibilities. If we think that all that's important about an agents beliefs is their attitudes towards the expressible propositions, then why not think that it is the expressions that are the objects of the agents beliefs?

So, this paper has two parts. Section 2 investigates how our framework of attitudes attached to sets of possible worlds works when there are beliefs about beliefs involved. Section 3 then investigates whether we can find a space of all possibilities once we want to account for beliefs about beliefs. Before moving to that discussion, we will briefly present some comments on what the question of choosing a universal space before we consider beliefs about beliefs.

1.1 Fixing states of nature

The starting point of this paper is that we have a fixed space of all the 'states of nature': $\mathcal{S}$. These are all the ways the world might be when beliefs are ignored. (This term comes from the economics and game theory literature.)

There have been some worries about how to understand this space of all possibilities. My own view is that we should take the possibilities be restricted kinds of objects. They won't fix absolutely everything about nature, but only those things that we're interested in, or aware of. We might have a collection of "all possibilities" which just contains two possibilities: one where the coin lands heads, and one where it does not land heads. If all we care about, or are aware of, is whether this coin lands heads, then this really does describe all the possibilities. Here we aren't *ruling out* any possibilities, we are simply dealing with a more coarse grained picture than '*really-all*' possibilities.

There are various ways to formally spell out the above idea. Here's one of them: start with a collection of basic propositions, which are not initially thought of as sets of possibilities but as linguistic, or pseudo-linguistic entities (let's call them sentences to distinguish them from the picture we're adopting of propositions simply as sets of possibilities). The possibilities are constructed from them as various ways those sentences might be true/false. We might thus think of this as something like a propositional logic valuation, with the initial propositions being the sentential variables. This is slightly different to Chalmers

(2011)'s construction which has the possibilities as truth assignments to whole languages, but then rules out some of them as impossible, for example those that have both φ and $\neg\varphi$ as true. I'm not sure yet how the two constructions differ and why one or the other might be preferable.

This sort of picture has payoff elsewhere: it allows us to account for Frege's puzzle whilst keeping propositions as collections of possible worlds by allowing our possibilities to be more flexible than 'real', or 'physical' or even 'logical' possibilities. In this setting we *can* have possible worlds where Superman is strong and Clark Kent is not strong, actually Chalmers (2011) points this out too. And perhaps my suggestion is very similar to his. We might also allow our possibilities to be governed by non-classical logics if we want and this will still naturally fit into the framework (Williams, 2018).

This sort of picture avoids one of the key worries that is given to accounts trying to specify all possible worlds. That worry is that there will be too many possibilities to be gathered together into a set. The argument has a picture of the possibilities being more like real possible worlds, and in particular they will specify how many objects exist. In my possibilities nothing may be said about exactly how many objects exist unless one initially plugs in sentences like 'there are exactly 3 objects'. Once one adopts the 'real-possibilities' perspective and thus needs to specify exactly how many objects there are, then for any cardinal κ, there should be a possibility where there's exactly κ-many objects (after all, we're not certain that that's not possible). But we've now got too many possibilities to form a *set*: there are at least as many possibilities as there are cardinal numbers. (But no such collection can itself have a cardinal number, so it cannot be a set.)

This isn't a complete no-go for the possibility-theorist, she might simply want to reform probability theory to allow that the collections of worlds be class-sized. Ω can now be a proper class (a term for collections that are too big to form sets[1]). And $\mathcal{F}$ is a *collection* of sub-collections of Ω (instead of a set of subsets of Ω). $\mathcal{F}$ may thus be a super-class: a class containing some proper classes. One might worry about paradoxes of class theory arising, but nothing like Russell's paradox follows from this set-up. We're not trying to take the class of *all* classes; that would be problematic for familiar Russell paradox like reasons. Instead we're just taking the class of sub-classes of a fixed class: Ω (or perhaps a subcollection of this maximal $\mathcal{F}$ as $\mathcal{P}(\Omega)$). Indeed, it will be coherent; if we think about it in set theoretic terms it'll look just like some layers above the set theoretic hierarchy that we accidentally forgot to call sets.

This sort of problem is only a worry for the view I take where the possibilities just specify truths of some basic-propositions if we take there to initially be proper-class-many basic-propositions of which we are aware. For example, for each cardinal κ we are aware of a proposition saying that there are exactly κ-many worlds. That is a lot of propositions to be aware of and have opinions about! Limited agents like me, certainly don't seem to have these conceptual

[1] Though many authors have argued about exactly how to understand classes; I'm taking a 'collective' notion rather than a 'logical' one.

resources. I don't know what all the cardinals are; I certainly don't have a name for each cardinal. However, even if one thinks there are class-many initial basic-propositions to be accounted for, this only leads us to the presentation of probability theory over classes, which I think is possible.

Hajek (ms) points out that this sort of account faces a different sort of problem. If our possibilities are restricted entities that only specify what happens with respect to various propositions we are aware of, it is now a legitimate question to ask what happens when we become aware of more basic-propositions. (Note: we are becoming aware of more *propositions* and not directly of more *possibilities*: it's not like we ruled out any possibilities before, our previous opinions were simply over a too coarse grained space.) This "update" of agents' beliefs cannot be done just by conditionalization since conditionalization only applies when we have a prior probability of that happening, and in these cases before becoming aware of the proposition we simply didn't have an opinion. But this problem looks like it's perpendicular to the question of what 'all the possibilities' are. Even if we fix the space of worlds as "really" all possibilities, we face the same challenge once we want to account for slightly limited agents who do not have attitudes towards *all* propositions, i.e. all collections of possibilities, but instead have a opinions in some limited collection of propositions. This is because we perhaps want our agent to start having more opinions over time. So even by fixing a collection of all possibilities we still have to account for updates, that look very similar to the challenge faced by the restriction of Ω. Of course, one might consider very idealised agents who have Ω as all possibilities and $\mathcal{F}$ as $\mathcal{P}(\Omega)$. But this is very idealised.

I agree that once one accepts the kind of picture of possibilities that I am proposing then there is pressure on us to develop a good account of how one 'learns' by becoming aware of more propositions. Some work has been done trying to address this question. One key proposal is 'Reverse Bayesianism' (e.g. Karni and Vierø, 2013; Bradley, 2017), which is roughly the idea that when you add new propositions this shouldn't directly change one's attitudes towards the old propositions. Steele and Stefansson (ms) put pressure on this idea giving examples where simply becoming aware of new propositions seems to allow for rational changes with respect to the old propositions. This puts pressure on our framework because it says that looking at a restricted collection of propositions might give misleading attitudes. (Their setup differs from mine in that they *are* ruling out possibilities in their setup, instead of my picture which merely considers coarse graining. So I still need to check that their arguments go through on my picture.)

It's interesting to note that I think we are on both horns of Hajek (ms)'s dilemma simultaneously. Restricting our possibilities doesn't immediately mean we avoid the too-many-worlds question; for that we have to argue that all these initial propositions specifying exactly how many worlds there are aren't things we have opinions about. And allowing the worlds to have *all* the worlds doesn't avoid the how-to-update-when-we-learn-the-unexpected worry unless we also assume that the agent has opinions about all propositions (ie. all collections of worlds), but that is very idealised. But I think both problems are overcome-able.

2 Interactive beliefs and sets of possibilities

We want to allow our agents to have beliefs about each others' beliefs. For example we want to allow:

- Rosy has beliefs about nature, e.g. coins,
- Ben has beliefs about Rosy's beliefs about coins,
- Rosy has beliefs about Ben's beliefs about Rosy's beliefs about coins,
- etc

Our agents have attitudes towards propositions. So in accounting for Ben's beliefs about Rosy's beliefs about coins we need there to be propositions about Rosy's beliefs. And propositions are taken as sets of possible worlds. So each world need to specify facts about beliefs as well as specifying facts about nature. But remember that we're understanding beliefs as defined *over* the space of possibilities. So our possibilities need to perform two roles: they need to determine what the agents beliefs are, or specify facts about the agents' beliefs, and they also need to constitute the space over which the agent's beliefs are defined. We thus have a circular specification of the possible worlds: what worlds there are depend on what the agents' beliefs might be like, which in turn depends on what worlds there are. This is what causes problems for there being a *universal* space of possible worlds. But it doesn't itself cause problems for providing models where the worlds perform both these roles.

I'll now fix a particular form of the agents attitudes. Actually I think this discussion is a general one that doesn't depend on the form of the agents beliefs: be they probabilities, full-beliefs, knowledge or whatever. But to have something concrete to work with I'll pin it down.

Suppose we have a fixed space of all the possible worlds: W. The agent has a certainty-set: a collection of worlds that *they* consider possible $\Omega \subseteq W$. The agent's credence function is then defined over Ω. We want to allow flexibility that the agent doesn't have to have beliefs in *all* propositions, instead they have an algebra of propositions $\mathcal{F} \subseteq \mathcal{P}(\Omega)$. They then have a credence function $p : \mathcal{F} \to \mathbb{R}$. This is the set-up used in Moss (2018). The inclusion of $\mathcal{F}$ and additional assumptions that we impose on it actually turns out to be important for universality results, although in this draft I don't discuss this too much.

So to give a possible worlds structure we need some worlds where each world specifies what beliefs the agent has in that world and what facts about nature hold. We thus want something with the following structure:

Definition 2.1. A *possible worlds structure* over a space of states of nature $\mathcal{S}$ with a collection of agents Agents is some

$$(W, V, \{(\Omega^a_w, \mathcal{F}^a_w, p^a_w) \mid a \in \mathsf{Agents}, w \in W\})$$

where

- W is some non-empty set, called 'the possible worlds', or 'the possibilities'.
- $V : W \to \mathcal{S}$ assigns to each world a state of nature.
- For each agent a, and world w, $(\Omega^a_w, \mathcal{F}^a_w, p^a_w)$ is a probability space with $\Omega^a_w \subseteq W$, i.e.:
  - $\Omega^a_w \subseteq W$,
  - $\mathcal{F}^a_w$ is a Boolean/σ-algebra over Ω^a_w,
  - $p^a_w : \mathcal{F}^a_w \to \mathbb{R}$.

It might be easier simply to see an example.
Consider the following description of a case:

> Rosy sees a fair coin tossed and doesn't know if Bob saw it or not.

We can create a model giving the possibilities that are appropriate for this case, and pictorially describe it as in fig. 1

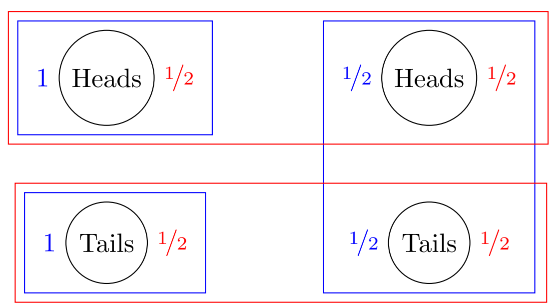

Figure 1: Example of a possible world structure. Rosy's beliefs are represented in red, Bob's are in blue.

These sorts of models are familiar in modal logic, they're roughly Kripke structures, or in this case a probabilistic variant of Kripke structures. They are closely related to type spaces, which are used in game theory.[2] The construction of such models by Harsanyi (1967) created relief for game theorists allowing work to be done on games with incomplete information breaking the seemingly problematic infinite regress.

So our first challenge has been overcome: we can capture interactive beliefs in this way where the propositions are understood as sets of possibilities.

The worlds have relationships to eachother, as governed by the agents opinions at the worlds. This is a consequence of adopting the picture of agents' attitudes attaching to propositions, which are defined as collections of worlds. Someone

[2] The difference being that in the type spaces framework the worlds are 'factorised' into their state-of-nature and interactive-beliefs components. The relationship between these two frameworks is discussed in Bjorndahl and Halpern (2017).

who takes the collection of all possibilities to be real, genuine possibilities doesn't avoid this point. And uniquely identifying worlds involves determining these relationships to other worlds. This causes challenges for determining a space of all possibilities because a specification of a possible world involves facts about other worlds.

3 Universal space

We have a fixed space of states of nature $\mathcal{S}$ and a collection of agents who are being represented. Given just this information we would like to have a space of all the possibilities that is fixed before we start thinking about the story going along with the case. This can be given by choosing a universal space.

Exactly what are we looking for from this universal space? The idea is that every possible combination of beliefs of our agents is represented in this universal space. But which combinations of beliefs are possible? Which beliefs are possible depends on which worlds are around, which depends on which beliefs are possible. This circular nature of the setup is what makes it complicated to determine all the possible worlds. But can it be done? To determine whether there is such a universal space we will investigate some options for what counts as capturing all possible beliefs about beliefs.

If we use the official definition of the agents beliefs as above we are left with little hope for finding a space with all the beliefs. Officially agent's attitudes are defined over sets of possible worlds, so if we change the possible worlds, we change the attitudes. Consider for example one belief structure with $W = \{1, 2, 3, 4\}$ and another with $W = \{a, b, c, d\}$ (remember we simply said W was a non-empty set, so either of these two are legitimate carrier sets). Officially any beliefs defined over these two spaces will simply be different. But they might nonetheless be 'essentially' the same. In particular we might have an *isomorphism* between the two possible world structures, for example they might both be equally well representable by our fig. 1.

So we have to be more subtle about exactly what we mean by all possible beliefs and what counts as equivalent structure.

A first attempt might be that every possible world structure can be embedded, via isomorphism, into the universal structure. This will similarly be too restrictive. Consider for example fig. 2:

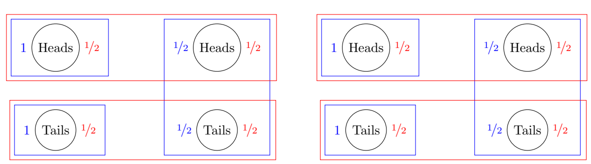

Figure 2: Two-copies of the structure

This doesn't *really* contain any additional beliefs. It just has multiple copies of the beliefs found in fig. 1.

The question of whether there exists a universal space has been well considered in the context of type spaces. Here's some motivation that is given by Heifetz and Samet for giving a universal space:

> It would, therefore, be nice to know that there is a universal knowledge space, in which every state of mind of the players in whatever knowledge space with the same states of nature is represented.... If such a big space exists, we could in principle always carry out the analysis in it, with no fear of neglecting any relevant state of affairs.Heifetz and Samet (1998a)

In that literature, the following criteria is used for a universal space: every space can be embedded with a "structure preserving map" into the universal space. It's not yet clear to me how this criteria is appropriate for the universal space to play the roles identified in the above quote. But for now we will proceed with it.

Exactly what the structure preserving map is depends on what structure is involved in our cases, ie. what kinds of attitudes are being considered. For our case as above, a structure preserving map would be defined as follows.

Definition 3.1. $f : \mathfrak{M} \to \mathfrak{M}^*$ is a structure preserving map '*morphism*' if:

- nature is the same in w and $f(w)$.
- $\Omega^*_{f(w)} = f[\Omega_w]$
- If $E \in \mathcal{F}^*_{f(w)}$ then $f^{-1}(E) \cap \Omega_w \in \mathcal{F}_w$.
- $m^*_{f(w)}(E) = m_w(f^{-1}(E) \cap \Omega_w)$ for all $E \in \mathcal{F}^*_w$.

$\mathfrak{M}^*$ is a *universal possible world structure* if for every $\mathfrak{M}$ there is a morphism from $\mathfrak{M}$ to $\mathfrak{M}^*$.

So for example we get that there is a morphism from the two-copies of the coin-possible-world-structure from fig. 2 to the single coin-possible-world-structure from fig. 1. Similarly we get a morphism from fig. 3 to fig. 1

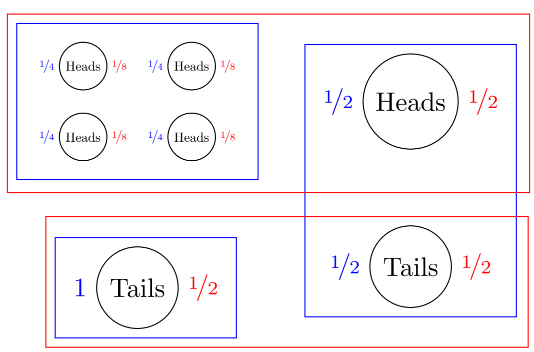

Figure 3: Multiplying worlds in structures

The existence or non-existence of a universal space, understood in this 'morphism' sense, has been well studied for various cases in the type-spaces literature. The existence of a universal type space was shown, with special assumptions, in Mertens and Zamir (1985) and a series of papers followed which weakened the assumptions needed, resulting in a rather general existence proof in Heifetz and Samet (1998b).[3] Those papers only focused on the probability component: there was no additional certainty component. Heifetz and Samet (1998a) shows that there is no universal *knowledge* space.

The key difference isn't the qualititive vs quantative nature. Instead it's that if we require that the probability measures are σ-additive then we have only finite "information carried" by the structure, we have:

$$p_w(E_1 \cup E_2 \cup \ldots) = \lim_{n \longrightarrow \infty} p_w(E_1 \cup \ldots \cup E_n)$$

But this isn't going to typically be the case for the knowledge component, which for our attitudes would be the agents' certainty set: Ω. Just knowing that $E_1 \not\supseteq \Omega$, $E_2 \not\supseteq \Omega, \ldots$, doesn't tell us whether $E_1 \cup E_2 \cup \ldots \supseteq \Omega$ or not. The 'certainty' of infinite unions isn't uniquely characterised by the certainty of the disjuncts.

If we only take our probability measures to be *finitely* additive, then we are in a situation analogous to the knowledge case. We similarly don't have that attitudes towards these infinite-union-propositions aren't uniquely characterised by her finite-level beliefs. And Meier (2006) shows we get an impossibility result. So we only get a universal space if we make rather strong assumptions on our agents' probabilities, and not at all if we have these qualitative components.

We will actually be able to recover the existence of a universal space without making such strong requirements on agents attitudes by restricting the kinds of propositions that the agents have attitudes towards, or expand our notion of 'equivalent beliefs' to just care about the agents' attitudes towards a restricted range of propositions. The problem was caused by attitudes towards infinite unions not being determined by attitudes towards the finite propositions. One might simply say that this is not a relevant or important component of one's belief state. We can do this by thinking about the *expressibility* of propositions in, e.g. propositional logic with operators for the agents attitudes, ie. some version of propositional modal logic. We start with facts about nature. We can also have propositions which are expressible by, e.g. $B(\text{Heads})$, and by $BB(\text{Heads})$ and so-on, but we don't care about one's attitudes towards the proposition which would be expressed by "$\neg B(\text{Heads}) \vee \neg BB(\text{Heads}) \vee \neg BBB(\text{Heads}) \vee \ldots$": this is an infinite disjunction and perhaps our agents simply don't have opinions regarding such propositions. So this picture is: an agents opinions are uniquely determined by their attitudes toward a limited range of *expressible* propositions.

We thus come to a construction that is familiar under a different guise: the canonical model construction from modal logic. These will ultimately

[3] See references in Heifetz and Samet (1998a) for a more detailed historical overview. The move to generality of these results successively dropped various assumptions on the structure of $\mathcal{S}$.

find something like a universal space. The claim is that the canonical model construction provides a universal space. I think we can link it to the above definition of a universal space by not allowing $\mathcal{F}$ to vary but instead fixing it as the propositions expressible by such finite sentences. But this still needs to be checked.

In fact Meier does something slightly different and he shows the existence of a universal space for finitely additive spaces. He shows that if we allow (rather than require) the measure space to simply be a σ-algebra then we get a universal space. The universal space will only have a σ-algebra, and the other spaces which may not still get mapped into it with a morphism which will ignore how it works on the non-measurable sets. I don't think this should count as a universal space unless we add a claim that the only thing that matters to the agents beliefs are her attitudes towards thus-and-so propositions. That's what we do by focusing on the move to expressibility.

A canonical model construction from modal logic defines a special kind of possible world structure $\mathfrak{M}^*$ where if φ is true in some world of some possible world structure then there is a world in the $\mathfrak{M}^*$ which also makes it true. It's constructed by defining a $\mathfrak{M}^*$ by taking the worlds to be satisfiable collections of sentences then recovering the other components of the structure from the sentences in the world. These can quite generally be found. I still need to check exactly when they can be defined, but it removes the worry that we had from above: there's now no non-existence results to be dealt with, at least none that I know of now.

4 Conclusion

So we can find models allowing us to have this simple picture of agents having attitudes towards propositions, conceived of as sets of possibilities, while allowing for interactive beliefs. Indeed this was the key insight from Harsanyi (1967) that allowed game theorists to study games with incomplete information.

We have also mentioned various possibility and impossibility results regarding finding a universal space. If we only allow for probabilities and we have tight constraints on the agents' beliefs, namely that they just have probabilities that are σ-additive, then there is an existence theorem. However once we relax this we need to be more careful. If we say that all that matters for identifying beliefs are her attitudes towards a limited range of expressible propositions, then we will be able to recover the existence of a universal space. But if one thinks that we need to account for her attitudes towards *all* propositions, then we have impossibility results. I don't yet know if I think attitudes towards expressible propositions are enough to uniquely identify one's beliefs, and thus whether we are facing a possibility or impossibility result.

This means, for example, that one cannot have attitudes towards propositions we might intuitively phrase as "Rosy is not fully introspectively certain that the coin lands heads". This refers to many levels of belief at once, and is not expressible in the formal logic system. So the restriction to propositions expressible is indeed rather restrictive. If we want to allow attitudes towards such "fully introspective" claims, we return to the realm of impossibility.

References

Adam Bjorndahl and Joseph Y Halpern. From type spaces to probability frames and back, via language. *arXiv preprint arXiv:1707.08738*, 2017.

Richard Bradley. *Decision theory with a human face*. Cambridge University Press, 2017.

David J. Chalmers. The nature of epistemic space. In Andy Egan and Brian Weatherson, editors, *Epistemic Modality*. Oxford University Press, 2011.

Al Hajek. Omega? Draft of 4th July 2018, ms.

John C. Harsanyi. Games with incomplete information played by bayesian players, I-III part I. The basic model. *Management science*, 14(3):159–182, 1967.

Aviad Heifetz and Dov Samet. Knowledge spaces with arbitrarily high rank. *Games and Economic Behavior*, 22(2):260–273, 1998a.

Aviad Heifetz and Dov Samet. Topology-free typology of beliefs. *journal of economic theory*, 82(2):324–341, 1998b.

Edi Karni and Marie-Louise Vierø. "Reverse bayesianism": A choice-based theory of growing awareness. *American Economic Review*, 103(7):2790–2810, 2013.

Martin Meier. Finitely additive beliefs and universal type spaces. *The Annals of Probability*, 34(1):386–422, 2006.

Jean-Francois Mertens and Shmuel Zamir. Formulation of bayesian analysis for games with incomplete information. *International Journal of Game Theory*, 14(1):1–29, 1985.

Sarah Moss. *Probabilistic knowledge*. Oxford University Press, 2018.

Katie Steele and H. Orri Stefansson. Belief revision for growing awareness. Draft of 4th July 2018, ms.

J Robert G Williams. Rational illogicality. *Australasian Journal of Philosophy*, 96(1):127–141, 2018.