

# Standard State Space Models of Unawareness (Extended Abstract)\*

Peter Fritz

Department of Philosophy,  
Classics, History of Arts and Ideas  
University of Oslo, Norway  
peter.fritz@ifikk.uio.no

Harvey Lederman

Department of Philosophy  
New York University, USA  
hs1306@nyu.edu

The impossibility theorem of Dekel, Lipman and Rustichini [8] has been thought to demonstrate that standard state-space models cannot be used to represent unawareness. We first show that Dekel, Lipman and Rustichini do not establish this claim. We then distinguish three notions of awareness, and argue that although one of them may not be adequately modeled using standard state spaces, there is no reason to think that standard state spaces cannot provide models of the other two notions. In fact, standard space models of these forms of awareness are attractively simple. They allow us to prove completeness and decidability results with ease, to carry over standard techniques from decision theory, and to add propositional quantifiers straightforwardly.

## 1 Introduction

Dekel, Lipman and Rustichini [8], hereafter, “DLR”, claim to show that “standard state space models preclude unawareness”. Their claim has achieved the status of orthodoxy.<sup>1</sup> The first task of this paper is to clear the way for standard state space models of unawareness by showing that the formal result DLR present does not establish their headline conclusion. DLR informally motivate certain axioms concerning unawareness, but in their formal impossibility result, they rely on the claim that these axioms hold at all states in the model. As section 2 argues, the assumption that axioms hold at all states of the model is unwarranted; in fact, DLR themselves reject it. While DLR’s formal results are valid, they are not sufficiently general to rule out standard state space models of unawareness. As we show, the impossibility results do not hold if one assumes only that DLR’s explicit assumptions about unawareness hold at some “real” states, as opposed to at all states. Even strengthening those explicit assumptions considerably does not reinstate the results.

But this does not yet vindicate standard state space models of unawareness. Section 3 presents a novel impossibility result which uses widely shared assumptions about awareness. The new impossibility theorem relies on the assumption that an agent who is aware of a conjunction is aware of its conjuncts. If awareness satisfies this assumption, then standard state space models do in fact preclude unawareness. We then distinguish three notions of awareness, and suggest that two important ones do not satisfy this assumption, leaving open the possibility that they could be adequately modeled by standard state space models.

The remainder of the paper continues in a more positive vein. We describe a simple class of standard state space models which represent key features of awareness. In section 4, we establish completeness and decidability for the logic of these models. We also show that adding propositional quantifiers, a

---

\*A draft of the full version of this paper is available at <http://users.ox.ac.uk/~hert2388/>

<sup>1</sup>See, e.g., [54, p. 2], [55], [24, p. 78], [25, p. 305], [26, p. 101], [40, p. 220], [31, p. 2790], [39, pp. 977–978], [27, p. 2454], [28, p. 257], [15, p. 42], [59, p. 516], and [58].

topic which has presented major difficulties for existing approaches to awareness, is straightforward in our standard state space models. In section 5, we present one way of implementing a choice-based approach to decision theory within these models, and show how non-trivial unawareness is consistent with speculative trade. Section 6 concludes.

## 2 DLR's Triviality Result

### 2.1 Standard State Space Models

Standard state space models for the knowledge and awareness of a single agent can be understood as certain tuples  $\langle \Omega, k, a \rangle$ .  $\Omega$  is required to be a set, called the set of *states*, from which a set of events is derived by taking an *event* to be a set of states.  $k$  and  $a$  are functions on events, which represent the agent's knowledge and awareness, respectively:  $k$  maps each event  $E$  to the event  $k(E)$  of the agent knowing  $E$ ;  $a$  similarly takes each  $E$  to the event  $a(E)$  of the agent being aware of  $E$ .

Such models are straightforwardly used to interpret a formal language in which one can talk about knowledge and awareness. Let  $L$  be such a language built up from proposition letters  $p, q, \dots$ , using a unary negation operator  $\neg$ , a binary conjunction connective  $\wedge$  and two unary operators  $K$  and  $A$ , respectively ascribing knowledge and awareness to the agent. Formulas of this language are interpreted relative to a model  $M = \langle \Omega, k, a \rangle$  and a valuation function  $v$  which maps each proposition letter  $p$  to the event  $v(p)$ . The interpretation uses a function  $\llbracket \cdot \rrbracket_{M,v}$  which maps each formula  $\varphi$  of  $L$  to the event expressed by  $\varphi$  in  $M$ , which can be understood as the set of states in which  $\varphi$  is true in  $M$ . To state the constraints on such a function let  $-E = \Omega \setminus E$ .

$$\begin{aligned} \llbracket p \rrbracket_{M,v} &= v(p) \\ \llbracket \neg \varphi \rrbracket_{M,v} &= -\llbracket \varphi \rrbracket_{M,v} \\ \llbracket \varphi \wedge \psi \rrbracket_{M,v} &= \llbracket \varphi \rrbracket_{M,v} \cap \llbracket \psi \rrbracket_{M,v} \\ \llbracket K\varphi \rrbracket_{M,v} &= k(\llbracket \varphi \rrbracket_{M,v}) \\ \llbracket A\varphi \rrbracket_{M,v} &= a(\llbracket \varphi \rrbracket_{M,v}) \end{aligned}$$

The agent being unaware of something can of course be understood as it not being the case that she is aware of it. We therefore syntactically use  $U\varphi$  as an abbreviation for  $\neg A\varphi$ . Similarly, we introduce the other connectives of classical propositional logic as abbreviations, using  $\vee$  for disjunction,  $\rightarrow$  for material implication,  $\leftrightarrow$  for bi-implication, and  $\top$  and  $\perp$  for an arbitrary tautology and contradiction, respectively. On the semantic side, we adopt the convention of writing  $fg$  for the composition of functions  $f$  and  $g$ , which allows us to write, e.g.,  $k \circ a(E)$  instead of  $k(a(E))$ .

In order to express general constraints on these models, we say that a formula  $\varphi$  is *valid on  $M$*  if  $\llbracket \varphi \rrbracket_{M,v} = \Omega$  for each valuation function  $v$ ; this can be understood as requiring  $\varphi$  to be true in every state of  $M$  according to every valuation function. In order to limit this constraint to a particular state  $\omega \in \Omega$ , we say that  $\varphi$  is *valid in  $\omega$*  if  $\omega \in \llbracket \varphi \rrbracket_{M,v}$  for each valuation function  $v$ .

These models count as “standard” in the sense of DLR. First, the events expressed by  $A\varphi$  and  $K\varphi$  are each a function of the event expressed by  $\varphi$ . (DLR call this “event-sufficiency”.) Second, negation is interpreted as set-complement and conjunction as intersection, so that all tautologies of classical propositional logic, such as  $p \vee \neg p$ , are interpreted as the set of all states in every model. (DLR call this assumption “real states”.)

## 2.2 DLR on Standard State Space Models

DLR introduce three constraints on awareness, which can be stated using the following three axioms:

Plausibility:  $Up \rightarrow (\neg Kp \wedge \neg K\neg Kp)$

$KU$ -Introspection:  $\neg KUp$

$AU$ -Introspection:  $Up \rightarrow UUp$

Their constraints on knowledge can be stated using the following three axioms:

Necessitation:  $K\top$

Monotonicity:  $K(p \wedge q) \rightarrow (Kp \wedge Kq)$

Weak Necessitation:  $Kp \rightarrow K\top$

Their main results are then:

**Theorem 1** (DLR). *Let  $M = \langle \Omega, k, a \rangle$  be a model on which Plausibility,  $KU$ -Introspection and  $AU$ -Introspection are valid.*

**1(i)** *If Necessitation is valid on  $M$ , then  $Ap$  is valid on  $M$ .*

**1(ii)** *If Monotonicity is valid on  $M$ , then  $Kp \rightarrow Aq$  is valid on  $M$ .*

**2** *If Weak Necessitation is valid on  $M$ , then  $Kp \rightarrow Aq$  and  $Ap \leftrightarrow Aq$  are valid on  $M$ .*

Our presentation of DLR's result differs in superficial respects from their original presentation. DLR do not present their constraints in terms of the validity of certain axioms. Thus, for example, instead of requiring  $KU$ -Introspection to be valid on  $M$ , they require  $k - a(E) = \emptyset$  for all events  $E$ . However, it is a routine exercise to show that this condition is equivalent to the validity of our corresponding axiom. The same point holds for the other axioms. In short, our later models will not be escaping their triviality result by a sleight of hand which depends on this presentation.

One reason for the variant presentation is that it will facilitate the later exposition. It also serves to demonstrate that standard state space models as discussed here are equivalent to what are now commonly known as neighborhood or Scott-Montague frames (see [56] and [44]). It is well known that given certain restrictions on the function interpreting knowledge, this function can be turned into a binary relation among states along the lines of those used by [37] and [29]. This representation as a binary relation is, in turn, formally interchangeable with the "possibility correspondences" introduced by [2] (see also [3]) and used throughout economic theory (see, e.g. [11]).

## 2.3 Two Kinds of States

In response to their triviality results, DLR suggest distinguishing informally between "real" states and "subjective" states. As we understand it, this distinction can be explained as follows. An epistemic model makes predictions about how an agent or group of agents will or would behave in particular situations. The model makes predictions about these situations by including states which represent them. The real states in a model are the states which represent situations the model is intended to describe. The model predicts an agent will behave a certain way in a particular situation just in case the agent behaves that way in the real state which represents that situation. The predictions of a single model are given by what holds at all its real states; the behavioral theory of a class of models is given by what holds at all real states in all its models.

A state in an epistemic model is subjective if it figures in the specification of what the agent knows or is aware of at some real state. According to this way of understanding real and subjective states, states may be both real and subjective. Suppose we wish to represent an agent who knows that a particular coin will be flipped, but who will not learn the outcome of this coin flip. If our model is intended to make predictions no matter how the coin lands, the subjective states needed to specify the agent's knowledge (heads and ignorant, tails and ignorant) will be exactly the real states; every state will be both real and subjective. But as DLR recognize, there is no reason to require all states to play both roles. In the earlier example, if we only wanted to make predictions about the situation in which the coin comes up heads, we would not count one of the states (tails and ignorant) as a real state; even so, to represent the agent's ignorance given heads, a state where the coin comes up tails would still have to be included as a subjective state. The point can also be illustrated with a less artificial example. Consider an analyst who wishes to model the interactions of agents who are rational, but who do not believe each other to be rational. To represent the beliefs of these agents, the analyst must include subjective states in which the agents are irrational. But although she includes these subjective states, the analyst has no intention of eliminating the claim that the agents are rational from the predictions of her theory. Rather, it is understood informally that these subjective states are not real; they do not represent situations the analyst aims to describe.

Put in the terms of our presentation, DLR's proposal is to allow subjective states in which the law of excluded middle,  $p \vee \neg p$ , may not be true. DLR have no intention of eliminating the law of excluded middle from the predictions of their theory. Rather, they introduce these subjective states to specify the agent's knowledge and unawareness at real states where classical logic holds. The theory of DLR's models is given by what holds at these classical real states, not by what holds at all states whatsoever. Still, since classical tautologies may fail in DLR's subjective states, their models violate the "real states" assumption, and so are not standard state space models.

But once we acknowledge that some states may not be intended to represent situations the analyst wishes to describe, a question arises: Why should one require DLR's three axioms on unawareness to be valid in *all* states? DLR do not argue for this assumption. At best, their arguments in favor of their axioms only motivate imposing these axioms at real states. These arguments provide no motivation for imposing the axioms on subjective states that are not real, since these states are merely included in the model to specify the agent's knowledge and unawareness at real states.

There is a general methodological principle in epistemic modeling that *axioms* are to be imposed at all states. But in the literature on awareness, following DLR, this methodological principle has long been abandoned. DLR's own non-standard models violate this requirement, as do the current leading proposals for representing awareness, for example that of [24]. The subjective states in DLR's models include states in which logical axioms, including the law of excluded middle, do not hold. In DLR's models, as in the ones we will propose, even logical axioms are allowed to fail at some states. The only difference between our proposal and theirs concerns which axioms are allowed to fail. DLR preserve their own axioms at all states, and move to non-standard models in which classical propositional logic may fail at subjective states. We will preserve classical propositional logic at all states, and work with standard models in which DLR's axioms may fail at subjective states.

Still, one might ask: How should we *understand* a state where DLR's axioms are false? DLR interpret their own subjective states as "descriptions of possibilities as perceived by the agent" (p. 171). This interpretation does not seem appropriate for our models, in which DLR's axioms may fail at subjective states. But such metaphorical interpretations of these states are unnecessary. Subjective states where the axioms of awareness are invalid are simply to be understood in terms of the agent's knowledge and awareness at real states where the axioms are valid.

In fact, we can give a direct argument for not imposing DLR's axioms at all states, and in particular,

for including states where *KU*-Introspection is invalid. First, by *AU*-Introspection, if an agent is unaware of  $p$ , then she must be unaware of being unaware of  $p$ . But then, by *Plausibility*, the agent does not know that she does not know that she is unaware of  $p$ . (Essentially, DLR already give this argument on p. 169.) In epistemic models, we generally represent an agent's not knowing  $q$  by including a state in which  $q$  is false. So, to allow for real states in which the agent does not know that she does not know that she is unaware of  $p$ , we must include subjective states in which the agent knows that she is unaware of  $p$ , and so violates *KU*-Introspection.<sup>2</sup>

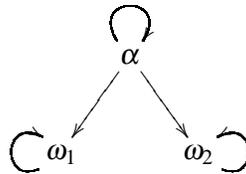
To sum up: after rejecting standard state space models, DLR propose that we should use models in which the laws of logic fail at subjective states. They implement this proposal by countenancing states where propositional logic fails, so that their models are non-standard. But if we allow models in which the law of excluded middle may fail at subjective states, we must also consider models in which other axioms, including DLR's, may fail at subjective states. DLR's formal results only apply to standard state space models in which their axioms are imposed at all states; the results do not concern standard state space models in which the axioms are imposed only at real states. As a consequence, these formal results cannot provide a basis for the conclusion DLR draw: that standard state space models preclude unawareness.

## 2.4 Non-Triviality

We have not argued against the validity of DLR's three axioms in real states – states representing the situations to be modeled. In our setting, we can formalize the distinction between real states and subjective states which are not real, by only assuming the validity of the axioms in some subset of the states in a model. We can then ask: is assuming the validity of the three axioms in such distinguished real states enough to lead to triviality? The following example shows that it is not:

**Theorem 2.** *There is a model  $M = \langle \Omega, k, a \rangle$ , state  $\alpha \in \Omega$  and event  $E \subseteq \Omega$  such that *Necessitation* is valid on  $M$ , *Plausibility*, *AU*-Introspection and *KU*-Introspection are valid in  $\alpha$ , and  $\alpha \notin a(E)$ .*

*Proof.* Let  $\Omega = \{\alpha, \omega_1, \omega_2\}$ . Define a binary (accessibility) relation  $R$  on  $\Omega$  as follows:



$R$  induces a possibility correspondence  $P$  such that  $P(\sigma) = \{\tau : R\sigma\tau\}$ . With  $P$ , define  $k$  and  $a$  such that for all  $F \subseteq \Omega$ :

$$k(F) = \{\sigma \in \Omega : P(\sigma) \subseteq F\}$$

$$a(F) = \begin{cases} \{\omega_2\} & \text{if } \omega_1 \in F \text{ and } \omega_2 \notin F \\ \Omega & \text{otherwise} \end{cases}$$

It is routine to check that  $M = \langle \Omega, k, a \rangle$ ,  $\alpha$  and  $E = \{\alpha, \omega_1\}$  witness the claim to be proven.<sup>3</sup> □

<sup>2</sup>This argument differs from DLR's main proof of triviality, since it only assumes that the axioms hold at the real state.

<sup>3</sup>As suggested in [43], one might consider an extension of *Plausibility* along the following lines: For each natural number  $n$ , let  $n$ -*Plausibility* be  $Up \rightarrow (\neg K)^n p$ , where  $(\neg K)^n \varphi$  is defined inductively by the two clauses  $(\neg K)^1 \varphi = \neg K \varphi$  and  $(\neg K)^{n+1} \varphi = \neg K(\neg K)^n \varphi$ . For each natural number  $n$ ,  $n$ -*Plausibility* is valid in  $\alpha$ .

This shows that DLR's Theorem 1(i) cannot be extended to standard state space models in which DLR's three axioms are only required to be valid in real states. In fact, the model used in the above proof of Theorem 2 can also be used to show that DLR's other two results cannot be extended either. For 1(ii), note that Monotonicity is valid on  $M$ , and that  $\alpha \in k(\Omega)$  although  $\alpha \notin a(E)$ . For 2, note first that Weak Necessitation is valid on  $M$ , and that  $\alpha \in a(\Omega)$  but as before  $\alpha \in k(\Omega)$  and  $\alpha \notin a(E)$ . More generally, any state in any model which satisfies both Necessitation and Monotonicity, in addition to DLR's three axioms, will be a counterexample not just to extensions of DLR's Theorem 1(i), but also to extensions of their Theorems 1(ii) and 2.

We conclude that none of DLR's three triviality results show that standard state space models preclude unawareness. One might wonder whether plausible strengthenings of the axioms on knowledge and unawareness allow us to reinstate the triviality results. In the full paper, we argue first that this cannot be achieved by strengthening their axioms governing knowledge, and, second, that it cannot be achieved by a particular strengthening of the axioms governing unawareness. The theorems and proofs may be found in Appendix A.

### 3 Three Kinds of Awareness

#### 3.1 A New Triviality Result

DLR's result had limited implications for state space models because it depended on the validity of their axioms at all states. Is there a triviality result which only uses the validity of axioms on awareness in real states, rather than their validity in all states? In fact, as we now show, widely accepted axioms on awareness *do* lead to triviality even if they are imposed only at real states. The result uses the following two axioms:

$$\text{AS: } A\neg p \rightarrow Ap$$

$$\text{AC: } A(p \wedge q) \rightarrow (Ap \wedge Aq)$$

Awareness is widely assumed to satisfy both of these axioms; see, e.g., [43, pp. 274–275], [18, p. 331] (axioms A1 and A2) and [25, p. 309] (axioms 1 and 2).

As the next theorem shows, these axioms lead straightforwardly to triviality.

**Theorem 3.** *Let  $M = \langle \Omega, k, a \rangle$  be a model and  $\alpha \in \Omega$  such that AS and AC are valid in  $\alpha$ . Then  $Ap \rightarrow Aq$  is valid in  $\alpha$ .*

*Proof.* Consider any events  $E$  and  $F$ , and assume  $\alpha \in a(E)$ . Since  $E = E \cap \Omega$ ,  $\alpha \in a(E \cap \Omega)$ , and so by AC,  $\alpha \in a(\Omega)$ . By AS,  $\alpha \in a(\emptyset)$ . Since  $\emptyset = \emptyset \cap F$ ,  $\alpha \in a(\emptyset \cap F)$ , and so by AC,  $\alpha \in a(F)$ .  $\square$

The crucial difference between this and DLR's triviality result is that AS and AC are only assumed to be valid in a distinguished state, for which it is shown that non-trivial unawareness in it is ruled out.

But does awareness really satisfy both AS and AC? In the following, we will focus in particular on AC, arguing that for some important notions of unawareness, being aware of a conjunction does not entail being aware of its conjuncts.

#### 3.2 Attending vs. Conceiving vs. Processing

In the literature on awareness, it is uncontentious that there is no single attitude of awareness; what is expressed by "aware" is a loose cluster of notions. This was noted at the very start of the literature, as

witnessed by the lengthy discussions in [10]; another detailed discussion can be found in [55]. We will argue that that at least some important notions of awareness do not satisfy AC (for others, as we will see, the situation is more complex).. In order to do so, we roughly distinguish the following three ways of understanding a claim of the form “The agent is aware of ...”:

- (i) The agent is attending to ...
- (ii) The agent has the conceptual resources required to conceive of ...
- (iii) The agent is able to process ...

We will introduce these notions – and distinguish between them – using various examples found in the literature.

Consider first attention. An influential example, which first appeared in [16], and which is discussed at length by DLR and numerous places in the subsequent literature, is based on the following quote from one of Arthur Conan Doyle’s Sherlock Holmes stories [9]:

“Is there any other point to which you would wish to draw my attention?”  
 ‘To the incident of the dog in the night-time.’  
 ‘The dog did nothing in the night-time.’  
 ‘That was the curious incident’ remarked Sherlock Holmes.”

Holmes’s interlocutor is Inspector Gregory, a Scotland Yard detective. Before Holmes pointed out to Gregory that the dog did nothing in the night-time, Gregory was *unaware* of the dog doing nothing in the night-time. Gregory’s state of unawareness is naturally understood as one of *inattention* – Holmes makes Gregory aware of the dog doing nothing in the night time in the sense of bringing this fact to his attention.

Gregory’s failing to attend to the dog doing nothing in the night-time must be sharply distinguished from Gregory’s not being able to conceive of the the dog doing nothing in the night-time. Before Holmes alerted Gregory to the dog doing nothing in the night-time, Gregory possessed the concepts required to entertain thoughts about the dog doing nothing in the night time. Contrast this with the following example for unawareness from [10, p. 40]:

“How can someone say that he knows or doesn’t know about  $p$  if  $p$  is a concept he is completely unaware of? One can imagine the puzzled frown of a Bantu tribesman’s face when asked if he knows that personal computer prices are going down!”

The relevant state of unawareness in this example is not merely a matter of the agent failing to attend to the relevant event or subject matter. For example, if one is unaware in the sense of being unable to conceive of an event, it must be that one does not understand the words for those notions in any language. Contrast this with the case of Inspector Gregory. Gregory understands what Holmes says: he can conceive of the dog’s doing nothing. But the purported example of inconceivability does not have this structure: the tribesman is supposed to be unable to think about computers using any of his conceptual resources, no matter what he attends to. The two notions of awareness – attending to versus being able to conceive of – are therefore clearly distinct.

The third notion of unawareness we want to single out is one which [10] (see also [19] and [20]) focus on; it can be understood as an attempt to deal with what is known as the “problem of logical omniscience” in epistemic logic. In standard state spaces, if two sentences  $\phi$  and  $\psi$  are equivalent in classical propositional logic, then  $K\phi$  and  $K\psi$  will be true in the same states. In particular, if  $K(p \vee \neg p)$  is true in a given state, then so is  $K\tau$  for any propositional tautology  $\tau$ . One of Fagin and Halpern’s reasons for developing a logic of awareness is to obtain logics which do not have this property. They write:

“The notion of awareness we use in this approach is open to a number of interpretations. One of them is that an agent is aware of a formula if he can compute whether or not it is true in a given situation within a certain time or space bound. This interpretation of awareness gives us a way of capturing resource-bounded reasoning in our model.”

Being unaware of  $\varphi$  in the sense of not being able to process  $\varphi$  is clearly distinct from failing to attend to  $\varphi$ : although Gregory did not attend to the dog doing nothing in the night-time, he had no difficulties processing the claim that the dog did nothing in the night-time. Not being able to process  $\varphi$  is also clearly distinct from not being able to conceive of  $\varphi$ : Gregory might not have been able to process an extremely complicated propositional tautology using only negation, conjunction and the sentence “the dog did nothing in the night-time”, but he clearly possessed all the concepts required to entertain it.

### 3.3 Awareness of Conjunctions

Let’s return to the new triviality result introduced at the start of this section. As already advertised, we believe that the principle AC, which says that an agent who is aware of a conjunction is aware of its conjuncts, may be plausible for one notion of awareness, but it is not for the other two.

Consider first awareness as the ability to process. This is plausibly a relation of agents to *sentences*, as part of what it takes to process ... is to be able to find out what the sentence “...” means. AC may hold if awareness is understood as the ability to process: It is natural to assume that an agent who is able to process a conjunction “... and —” is also able to process “...” and “—”. As noted already in [10, p. 54], even this may fail: an agent might be able to recognize that a very long sentence has the form  $\varphi \wedge \neg\varphi$ , and so be able to process it, although she is unable to process the complex  $\varphi$  on its own. Resolving this controversy may require distinguishing further among different notions of processing, and the appropriate resolution may depend on the intended application.

The relations of attention and conceivability are different from the ability to process. In particular, they are plausibly relations agents have not to sentences but to what sentences express. We might call these entities *contents* or *propositions*, but in keeping with the terminology employed above, we call them *events*. In the full paper, we discuss the difference between sentences and events in more detail, and motivate the assumption of a *coarse-grained theory of events*, according to which events form a complete atomic Boolean algebra. A consequence of this assumption – which is implicit in all standard state-space based modeling techniques – is that sentences which are equivalent in propositional logic have identical contents.

With this understanding of attention and conceivability as relations of agents to coarse-grained events, consider first attention. Assume that after the conversation with Holmes quoted above, Gregory is alone thinking about the case, and attending to the event of the dog barking in the night ( $p$ ). He is not, however, attending to event of Holmes at that moment smoking a pipe ( $q$ ). It is then natural to say that Gregory is also not attending to the conditional event that *if* Holmes is currently smoking a pipe *then* the dog barked in the night ( $q \rightarrow p$ ). But notice that according to the coarse-grained Boolean theory of events, the event that the dog barked in the night ( $p$ ) is identical to the event that if Holmes is smoking a pipe then the dog barked in the night, and the dog barked in the night  $((q \rightarrow p) \wedge p)$ . So if AC were valid, then since Gregory is attending to the event of the dog barking in the night, he would be attending to the event that if Holmes is smoking a pipe then the dog barked in the night. But by assumption Gregory is not attending to this last event. Thus it follows from the coarse-grained theory of events that AC must be rejected.

A similar example can be given if we understand awareness as conceivability. Assume our agent does not have the conceptual resources to entertain the event of there being a black hole. According to

the assumed coarse-grained theory of content, the event of there being a black hole and there being no black hole is identical to any event expressed by a propositional contradiction, such as the event of there being a sheep and there being no sheep. The agent might well have the conceptual resources to entertain the event of there being a sheep and there being no sheep, without having the conceptual resources to entertain the event of there being a black hole.

If we adopt, as usual, a theory of events which identifies the event expressed by sentences which are equivalent in propositional logic, AC appears to be inappropriate. Thus the new triviality result with which we started this section also does not establish that standard state space models preclude unawareness understood as inattention or inability to conceive.

## 4 Partitional Models

So far, we have shown that standard state-space models escape certain putative impossibility results for models of attention and conceivability. But this does not establish that standard space models can provide fertile models of these notions. In the remainder of the paper, we define, motivate, and examine a class of standard state space models for representing attention and the ability to conceive.

To show that our models generalize smoothly to the multi-agent case, from now on we use a language  $L_I$  parametrized to an arbitrary set of agent-indices  $I$  which is defined as the language  $L$  above, except that the operators  $A_i$  and  $K_i$  are indexed to  $i \in I$ . Models are consequently tuples of the form  $\langle \Omega, k^i, a^i \rangle_{i \in I}$ .

The models we will be working with are defined as follows:

**Definition 1.**  $\langle \Omega, R^i, \approx^i \rangle_{i \in I}$  is a partitional model if  $\Omega$  is a set and for each  $i \in I$ ,  $R^i$  is a binary relation on  $\Omega$  which is reflexive and transitive, and  $\approx^i$  a function which maps each  $\omega \in \Omega$  to an equivalence relation  $\approx^i_\omega$  on  $\Omega$ .

Here and in what follows, we make use of the fact that each equivalence relation corresponds to a unique partition, and *vice versa*; accordingly, we treat them as interchangeable.

Partitional models can be used to generate standard models in the following way:  $R^i$  specifies states of knowledge just as in Theorem 2. The idea behind  $\approx^i$  is that the events the agent is aware of at  $\omega$  are the events which are unions of sets of equivalence classes of  $\approx^i_\omega$  (equivalently: unions of sets of cells of the induced partition). So for each  $i \in I$ , let  $R^i$  and  $\approx^i$  determine functions  $k^i$  and  $a^i$  on events on  $\Omega$  as follows:

$$k^i(E) = \{ \sigma \in \Omega : P^i(\sigma) \subseteq E \}, \text{ where } P^i(\sigma) = \{ \tau : R^i \sigma \tau \}$$

$$a^i(E) = \{ \sigma \in \Omega : \text{for all } \rho \text{ and } \tau \text{ such that } \rho \approx^i_\sigma \tau, \rho \in E \text{ iff } \tau \in E \}$$

Let the standard model determined by a partitional model  $\langle \Omega, R^i, \approx^i \rangle_{i \in I}$  be  $\langle \Omega, k^i, a^i \rangle_{i \in I}$ , with  $k^i$  and  $a^i$  as just defined. On such a standard model,  $L_I$  can be interpreted as above; obviously, this induces a way of interpreting  $L_I$  directly on partitional models.

### 4.1 The Attitude of Attention

In order to motivate partitional models as models of limited attention, we suggest that attention in the sense we have been using the term should primarily be understood as an attitude towards questions. There are many available formal approaches to modeling questions (for an overview, see [35]). For concreteness, we'll adopt a standard approach, representing questions as partitions of the state space (see [17], building on [23] and [32]). Although we think the attitude to questions is primary, we will follow the literature on awareness, in axiomatizing a notion of attention which has events as its objects. The

relationship between this attitude to events and the attitude toward questions will be as follows: an agent attends to the question  $Q$  if and only if the agent attends to every partial answer to  $Q$ . Using partitions to model questions, partial answers are unions of sets of cells, corresponding to how standard models are derived from partitional models.<sup>4</sup>

## 4.2 Partitions for Conceivability

To motivate the use of partitional models of conceivability, assume that the agents to be modeled have the concept of negation and (infinitary) conjunction, so that the set of events they can conceive of are closed under complement and arbitrary intersection. This is mathematically equivalent to requiring that this set is derived from an equivalence relation as above.

## 4.3 An Example

It will be useful to have a concrete partitional model before us, as a running example. The following model shows that there are non-trivial partitional models; for simplicity, a single-agent case is specified. Let  $M = \langle \Omega, \mathcal{R}, \approx \rangle$ , with  $\omega R v$  iff  $\omega = 1$  or  $\omega = v$ , and  $\approx$  given by the following equivalence classes:

$$\approx_1: \{1\}, \{2, 3, 4\}$$

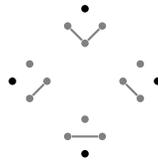
$$\approx_2: \{1\}, \{2\}, \{3, 4\}$$

$$\approx_3: \{1\}, \{3\}, \{2, 4\}$$

$$\approx_4: \{1\}, \{4\}, \{2, 3\}$$

Thus, at 1, the strongest event known by the agent is  $\Omega$ , and at each other state  $n$ , it is  $\{n\}$ . At each state  $n$ , the events the agent is aware of are the events which don't distinguish between any states in  $\Omega \setminus \{1, n\}$ .

Drawing the four states in a circle, starting with 1 at the top and going clockwise, we can draw each equivalence relation in a similar smaller circle, connecting two states by a sequence of lines if they are related by the relevant equivalence relation:



This is a partitional model in which there is non-trivial unawareness at each state. We will appeal to it below in order to show the consistency of various constraints.

## 4.4 Axioms

Given a class of models  $\mathcal{C}$ , a set of sentences  $\Sigma \subseteq L_I$  is the logic of  $\mathcal{C}$  if and only if  $\Sigma$  contains exactly those sentences which are valid on  $\mathcal{C}$ . Characterizing the logic of a class of models gives us a formal perspective from which to assess what assumptions our models encode about agents knowledge and awareness.

Thus we may ask: What is the logic of partitional models? Standard techniques on completeness results in modal logic are easily adapted to obtain the following result.

<sup>4</sup>Section 5 discusses related models developed in [33, 34].

**Theorem 4.** *A formula is valid on all partitional models if and only if it is derivable in the calculus with the following axiom schemas and rules:*

*PL: Any substitution instance of a theorem of propositional logic.*

*K-K:  $K_i(\phi \rightarrow \psi) \rightarrow (K_i\phi \rightarrow K_i\psi)$*

*K-T:  $K_i\phi \rightarrow \phi$*

*K-4:  $K_i\phi \rightarrow K_iK_i\phi$*

*A-Neg:  $A_i\phi \rightarrow A_i\neg\phi$*

*A-M:  $(A_i\phi \wedge A_i\psi) \rightarrow A_i(\phi \wedge \psi)$*

*A-N:  $A_i\top$*

*MP: From  $\vdash \phi$  and  $\vdash \phi \rightarrow \psi$  infer  $\vdash \psi$*

*K-RN: From  $\vdash \phi$  infer  $\vdash K_i\phi$*

*A-RE: From  $\vdash \phi \leftrightarrow \psi$  infer  $\vdash A_i\phi \leftrightarrow A_i\psi$*

*Moreover, the logic is decidable.*

A proof is given in Appendix B.

## 4.5 DLR Once More

Consider again DLR's three axioms. Given our discussion above, it is natural to consider partitional models where DLR's axioms are required to be valid in some distinguished state:

**Definition 2.**  *$\langle \Omega, \alpha, R^i, \approx^i \rangle_{i \in I}$  is a partitional DLR model if  $\langle \Omega, R^i, \approx^i \rangle_{i \in I}$  is a partitional model and Plausibility, KU-Introspection and AU-Introspection (for each  $i \in I$ ) are valid in  $\alpha$ .*

We now show that DLR's triviality result cannot be revived in partitional models:

**Theorem 5.** *There is a partitional DLR model  $\langle \Omega, \alpha, R^i, \approx^i \rangle_{i \in I}$  and an event  $E \subseteq \Omega$  such that  $\alpha \in U(E)$ .*

*Proof.* Simply distinguish state 1 in the model presented in section 4.3. □

We conjecture that the logic of partitional DLR models can be axiomatized as follows:

**Conjecture 1.** *Add the following axioms to the theorems of the axiom system in Theorem 4 and close under modus ponens:*

*P:  $U_i\phi \rightarrow (\neg K_i\phi \wedge \neg K_i\neg K_i\phi)$*

*AU:  $U_i\phi \rightarrow U_iU_i\phi$*

*A formula is derivable in this calculus if and only if it is valid in every distinguished state of every partitional DLR model.*

Note that  $\neg KU\phi$  can be derived using P, AU and K-T.

The present result shows that we can impose the DLR axioms without trivializing partitional models. But we confess to doubts about whether these axioms are appropriate. Just as with AC, once we understand more clearly the character of attention and conceivability, as well as the distinction between sentences and what they express, DLR's axioms become much less compelling. The clearest case concerns Plausibility and attention. A consequence of the contrapositive of Plausibility is  $K\phi \rightarrow A\phi$ . But this principle is false for attention. You know that there are more than four stars in the universe, but we doubt that you were attending to the question of how many stars there are prior to reading the previous clause. As we discuss in more detail in the full paper, the coarse-grained conception of content together with clarity about the notion of awareness to be modeled cast doubt on DLR's axioms.

## 4.6 Propositional Quantification

A challenge to some approaches to unawareness is to represent propositionally quantified statements. E.g., earlier models by Halpern made the claim that the agent knew she was unaware of something unsatisfiable (cf. [21] and [22]). In standard state space models such as ours, it is trivial to add propositional quantifiers without any such consequences. To do so, we write  $v[E/p]$  for the valuation function which maps  $p$  to  $E$  and every other proposition letter  $q$  to  $v(q)$ :<sup>5</sup>

$$\llbracket \forall p \varphi \rrbracket_{M,v} = \bigcap_{E \subseteq \Omega} \llbracket \varphi \rrbracket_{M,v[E/p]}$$

To illustrate that these quantifiers behave just as one would expect, note that in state 1 of the example described in section 4.3, the agent knows that she is unaware of something without there being something that she knows to be unaware of:  $K \exists p U p \wedge \neg \exists p K U p$  is true in this state.

## 4.7 Closure and Automorphisms

In partitional models, what agents are aware of (attend to/can conceive) is closed under negation and conjunction. One might wonder whether we can also impose the constraints that what agents are aware of must be closed under awareness and knowledge. In other words, whether there are models on which the following axioms are valid:

$$\text{A-4ij } A_i p \rightarrow A_i A_j p$$

$$\text{AK-4 } A_i p \rightarrow A_i K_j p$$

To provide models which validate these principles we adapt the coherence constraint of [13].<sup>6</sup> The idea behind it is most easily described for awareness as conceivability, taking the equivalence relations of partitional models to represent a relation of indistinguishability using conceptual resources available to the relevant agent at the relevant state. Coherence requires that if two states are indistinguishable in this way, then there must be a way of permuting the state space in a way which preserves all structural facts about knowledge and awareness, as well as all the events which the relevant agent is aware of at the relevant state.

Let  $M = \langle \Omega, R^i, \approx^i \rangle_{i \in I}$  be a partitional model. A permutation  $f$  of  $\Omega$  is an automorphism of  $M$  if for all  $i \in I$ ,

(i) for all  $x, y, z \in \Omega$ ,  $y \approx_x^i z$  iff  $f(y) \approx_{f(x)}^i f(z)$ , and

(ii) for all  $x, y \in \Omega$ ,  $R^i x y$  iff  $R^i f(x) f(y)$ .

A state  $x \in \Omega$  *coheres* if for all  $i \in I$  and  $y, z \in \Omega$  such that  $y \approx_x^i z$  there is an automorphism  $f$  of  $M$  such that  $f(y) = z$  and  $f \subseteq \approx_x^i$  (i.e.,  $\omega \approx_x^i \omega$  for all  $\omega \in \Omega$ ). It's routine to verify that A-4ij and AK-4 are valid in any coherent state of a partitional model.

Once again, the model presented in section 4.3 demonstrates the satisfiability of this constraint: every state in this model is coherent. Since the model also satisfies the DLR axioms at state 1, it shows that even if we were to uphold the DLR axioms, imposing them together with coherence would not trivialize state space models of awareness.

<sup>5</sup>See already [36], and [12] for a more systematic development. See [14] for results on the complexity of propositional quantifiers in the related setting of [13].

<sup>6</sup>The following notion of coherence differs importantly from that of [13] in that  $\approx_x^i$  here need not relate  $x$  only to  $x$ .

## 5 Decision Theory

In section 4.5 the example of the number of stars illustrated how one may believe and know things to which one is not attending; clearly this kind of inattention may also affect choice-behavior. One advantage of standard state spaces is that we can use the usual decision-theoretic framework to represent the effects of inattention on choice-behavior.<sup>7</sup>

The usual decision theoretic representation of an agent's beliefs uses a measure-space  $\langle S, \mathcal{B}, \mu \rangle$ .<sup>8</sup> To generate a partitional model, we enrich this description of the agent by selecting  $\mathcal{B}^C$ , a complete atomic subalgebra of  $\mathcal{B}$ , to generate a representation of what the agent attends to in the context of choice:  $\langle S, \mathcal{B}, \mu, \mathcal{B}^C \rangle$ . The atoms of  $\mathcal{B}^C$  are a partition of  $S$ , so this structure gives rise to a partitional model of unawareness. The distribution the agent “uses” in a choice context is given by letting  $\mu^C(E) = \mu(E)$  for all  $E \in \mathcal{B}^C$  and undefined otherwise. The events the agent “explicitly believes” in the context can then be defined as the events of which the agent is certain in  $\mu^C$ . The algebra  $\mathcal{B}^C$  can also be used to parametrize “expanding” and more generally “changing” awareness, represented as transitions between different complete complete atomic sub-algebras of  $\mathcal{B}$ .<sup>9</sup> Since different algebras will determine different *explicit* beliefs in different contexts, this changing awareness can also represent effects of limited attention such as framing effects or failures of recall.

An approach along these lines has already proven fruitful in epistemic game theory. [33] and [34] develop Harsanyi type spaces in which players' beliefs may be defined on different  $\sigma$ -algebras. If the algebras are taken as the events the agent is attending to, one may interpret these models as examples of agents who fail to attend to questions about the higher-order beliefs of others, and thus do not have *explicit* beliefs over events which can be defined only by the level- $n$  beliefs of others for large enough  $n$ .

### 5.1 Speculative Trade

An important test of approaches to unawareness has been how they fare with speculative trade ([26]). Building on the work of [2], [41] proved their famous “no-trade” theorem, illustrating the extreme strength of  $S5$  knowledge together with a common prior. One aim of representing “bounded” agents such as those with limited attention is to escape such paradoxes (for this perspective, see [45], [38]). Accordingly, we now provide a partitional DLR model with a common prior in which speculative trade is possible.

As is well known (see [16], [50], [49]), the “no-trade” theorem does not hold in general if agents' accessibility relations  $R^i$  are merely transitive and reflexive, but are not required to form an equivalence relation. Plausibility is incompatible with the  $R^i$  forming an equivalence relation ([42]). Still, the DLR axioms together with partitional awareness models impose substantial further constraints, which might be thought to rule out speculative trade. We now construct a partitional DLR model to show that speculative trade can still occur in the presence of DLR's axioms and nontrivial unawareness.

Let the states be  $W = \{1, 2, 3, 4, 5\}$  and the agents be Alice,  $A$ , and Bob  $B$ . The accessibility relations are defined so that:  $1R^A x$  iff  $x \leq 3$ ;  $5R^A x$  iff  $x \geq 3$  and otherwise  $wR^A x$  iff  $w = x$ , while  $R^B = W \times W$ . The partitions of the agents are induced by  $\approx_1^A = \approx_2^A = \{\{1\}, \{2, 3\}, \{4\}, \{5\}\}$ ;  $\approx_4^A = \approx_5^A = \{\{1\}, \{2\}, \{3, 4\}, \{5\}\}$ ; and for all  $w$ ,  $\approx_w^B = \approx_3^B = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}\}$ . The agents' common prior is the uniform one, and two acts  $f$  and  $g$  have utility as follows:  $f(1) = f(5) = 1$ ,  $f(2) = f(4) = 5$ ,  $f(3) = 7$ ;  $g(w) = 4$

<sup>7</sup>We do not here attempt to back-form what the agent is aware of from her choice-dispositions, as [46, 47] do for belief, and [52, 53] do for awareness.

<sup>8</sup>This can be derived in any of the standard ways: e.g. [51], [48], [1], [5], [30], [6].

<sup>9</sup>See the full paper for an alternative related to that of [31].

for all  $w \in W$ . If the agents update by conditionalization on their implicit knowledge, then Alice will invariably maximize utility by choosing  $f$  (since in states 2, 3, 4 she is certain it does better, and in states 1 and 5 she expects to gain  $1/3 \cdot 1 + 1/3 \cdot 5 + 1/3 \cdot 7 > 4$ ). Bob meanwhile does not update at all, so that he strictly prefers  $g$  (since  $4 > 2/5 \cdot 1 + 2/5 \cdot 5 + 1/5 \cdot 7$ ).

## 6 Conclusion

Standard state space models of attention and conceivability are at least as successful as current non-standard state space models. The non-standard models are, however, more complicated, and it is unclear that this complexity affords any advantages in predictive strength or accuracy. Standard state space models of these phenomena promise to lead to a rich and rewarding theory, posing technical and conceptual challenges, and offering connections to related work by linguists, philosophers and logicians – as well as work on bounded reasoning elsewhere in economic theory.

## 7 Acknowledgments

Thanks to Eddie Dekel, Bart Lipman and Tim Williamson for very helpful discussions of drafts of this paper. Peter Fritz was supported by an AHRC/Scatcherd European Scholarship and the German Academic Exchange Service; Harvey Lederman acknowledges the support of the Clarendon Fund.

## References

- [1] Francis J. Anscombe & Robert J. Aumann (1963): *A definition of subjective probability*. *Annals of mathematical statistics*, pp. 199–205, doi:10.1214/aoms/1177704255.
- [2] Robert J. Aumann (1976): *Agreeing to Disagree*. *The Annals of Statistics* 4(6), pp. 1236–1239, doi:10.1214/aos/1176343654.
- [3] Robert J. Aumann (1999): *Interactive epistemology I: Knowledge*. *International Journal of Game Theory* 28, pp. 263–300, doi:10.1007/s001820050111.
- [4] Patrick Blackburn, Maarten de Rijke & Yde Venema (2001): *Modal Logic*. Cambridge University Press, doi:10.1017/CB09781107050884.
- [5] Ethan D. Bolker (1967): *A simultaneous axiomatization of utility and subjective probability*. *Philosophy of Science*, pp. 333–340, doi:10.1086/288171.
- [6] John Broome (1990): *Bolker-Jeffrey expected utility theory and axiomatic utilitarianism*. *The Review of Economic Studies* 57(3), pp. 477–502, doi:10.2307/2298025.
- [7] Brian F. Chellas (1980): *Modal Logic: An Introduction*. Cambridge: Cambridge University Press, doi:10.1017/CB09780511621192.
- [8] Eddie Dekel, Barton L. Lipman & Aldo Rustichini (1998): *Standard state-space models preclude unawareness*. *Econometrica* 66(1), pp. 159–73, doi:10.2307/2998545.
- [9] Arthur Conan Doyle (1901): *The Adventure of Silver Blaze*. *The Strand Magazine*.
- [10] Ronald Fagin & Joseph Y. Halpern (1988): *Belief, awareness, and limited reasoning*. *Artificial Intelligence* 34, pp. 39–76, doi:10.1016/0004-3702(87)90003-8.
- [11] Ronald Fagin, Joseph Y. Halpern, Yoram Moses & Moshe Y. Vardi (1995): *Reasoning about Knowledge*. MIT Press.
- [12] Kit Fine (1970): *Propositional quantifiers in modal logic*. *Theoria* 36, pp. 336–346, doi:10.1111/j.1755-2567.1970.tb00432.x.

- [13] Peter Fritz (forthcoming): *Propositional Contingentism*. Forthcoming in *The Review of Symbolic Logic*.
- [14] Peter Fritz (unpublished): *Logics for Propositional Contingentism*.
- [15] Spyros Galanis (2013): *Unawareness of theorems*. *Economic Theory* 52, pp. 41–73, doi:10.1007/s00199-011-0683-x.
- [16] John Geanakoplos (1989): *Game theory without partitions, and applications to speculation and consensus*. Technical Report, Cowles Foundation Discussion Paper.
- [17] Jeroen Groenendijk & Martin Stokhof (1984): *Studies on the Semantics of Questions and the Pragmatics of Answers*. Ph.D. thesis, University of Amsterdam.
- [18] Joseph Y. Halpern (2001): *Alternative semantics for unawareness*. *Games and Economic Behavior* 37(2), pp. 321–339, doi:10.1006/game.2000.0832.
- [19] Joseph Y. Halpern, Yoram Moses & Moshe Y. Vardi (1994): *Algorithmic knowledge*. In: *Proceedings of the 5th conference on Theoretical aspects of reasoning about knowledge*, Morgan Kaufmann Publishers Inc., pp. 255–266, doi:10.1016/B978-1-4832-1453-5.50022-2.
- [20] Joseph Y. Halpern & Riccardo Pucella (2011): *Dealing with logical omniscience: Expressiveness and pragmatics*. *Artificial Intelligence* 175(1), pp. 220–235, doi:10.1016/j.artint.2010.04.009.
- [21] Joseph Y. Halpern & Leandro C. Rêgo (2009): *Reasoning about knowledge of unawareness*. *Games and Economic Behavior* 67(2), pp. 503–525, doi:10.1016/j.geb.2009.02.001.
- [22] Joseph Y. Halpern & Leandro C. Rêgo (2013): *Reasoning about knowledge of unawareness revisited*. *Mathematical Social Sciences* 65(2), pp. 73–84, doi:10.1016/j.mathsocsci.2012.08.003.
- [23] C. L. Hamblin (1973): *Questions in Montague English*. *Foundations of Language* 10, pp. 41–53.
- [24] Aviad Heifetz, Martin Meier & Burkhard C. Schipper (2006): *Interactive unawareness*. *Journal of Economic Theory* 130(1), pp. 78–94, doi:10.1016/j.jet.2005.02.007.
- [25] Aviad Heifetz, Martin Meier & Burkhard C. Schipper (2008): *A canonical model for interactive unawareness*. *Games and Economic Behavior* 62, pp. 304–324, doi:10.1016/j.geb.2007.07.003.
- [26] Aviad Heifetz, Martin Meier & Burkhard C. Schipper (2013): *Unawareness, beliefs, and speculative trade*. *Games and Economic Behavior* 77, pp. 100–121, doi:10.1016/j.geb.2012.09.003.
- [27] Sander Heinsalu (2012): *Equivalence of the information structure with unawareness to the logic of awareness*. *Journal of Economic Theory* 147(6), pp. 2453–2468, doi:10.1016/j.jet.2012.05.010.
- [28] Sander Heinsalu (2014): *Universal type structures with unawareness*. *Games and Economic Behavior* 83, pp. 255–266, doi:10.1016/j.geb.2013.11.012.
- [29] Jaako Hintikka (1962): *Knowledge and Belief*. Cornell University Press.
- [30] Richard C. Jeffrey (1983): *The logic of decision*, second edition. Chicago: University of Chicago Press.
- [31] Edi Karni & Marie-Louise Vierø (2013): “Reverse Bayesianism”: A choice-based theory of growing awareness. *The American Economic Review* 103(7), pp. 2790–2810, doi:10.1257/aer.103.7.2790.
- [32] Lauri Karttunen (1977): *Syntax and semantics of questions*. *Linguistics and Philosophy* 1, pp. 3–44, doi:10.1007/BF00351935.
- [33] Willemien Kets (2014): *Bounded Reasoning and Higher-Order Uncertainty*. Submitted MS.
- [34] Willemien Kets (2014): *Finite Depth of Reasoning and Equilibrium Play in Games with Incomplete Information*. Submitted MS.
- [35] Manfred Krifka (2011): *Questions*. In Klaus von Heusinger, Claudia Maienborn & Paul Portner, editors: *Semantics. An International Handbook of Natural Language Meaning*, 2, Berlin: Mouton de Gruyter, pp. 1742–1785.
- [36] Saul A. Kripke (1959): *A Completeness Theorem in Modal Logic*. *Journal of Symbolic Logic* 24, pp. 1–14, doi:10.2307/2964568.

- [37] Saul A. Kripke (1963): *Semantical Analysis of Modal Logic I: Normal Modal Propositional Calculi*. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* 9, pp. 67–96, doi:10.1002/malq.19630090502.
- [38] Harvey Lederman (2015): *People with Common Priors Can Agree to Disagree*. *The Review of Symbolic Logic* 8, pp. 11–45, doi:10.1017/S1755020314000380.
- [39] Jing Li (2009): *Information structures with unawareness*. *Journal of Economic Theory* 144(3), pp. 977–993, doi:10.1016/j.jet.2008.10.001.
- [40] Martin Meier & Burkhard C. Schipper (2014): *Bayesian games with unawareness and unawareness perfection*. *Economic Theory* 56, pp. 219–249, doi:10.1007/s00199-013-0788-5.
- [41] Paul Milgrom & Nancy Stokey (1982): *Information, trade and common knowledge*. *Journal of Economic Theory* 26(1), pp. 17–27, doi:10.1016/0022-0531(82)90046-1.
- [42] Salvatore Modica & Aldo Rustichini (1994): *Awareness and partitional information structures*. *Theory and Decision* 37(1), pp. 107–124, doi:10.1007/BF01079207.
- [43] Salvatore Modica & Aldo Rustichini (1999): *Unawareness and partitional information structures*. *Games and Economic Behavior* 27(2), pp. 265–298, doi:10.1006/game.1998.0666.
- [44] Richard Montague (1970): *Universal Grammar*. *Theoria* 36, pp. 373–398, doi:10.1111/j.1755-2567.1970.tb00434.x.
- [45] Stephen Morris (1995): *The Common Prior Assumption in Economic Theory*. *Economics and Philosophy* 11(2), pp. 227–253, doi:10.1017/S0266267100003382.
- [46] Stephen Morris (1996): *The logic of belief and belief change: A decision theoretic approach*. *Journal of Economic Theory* 69(1), pp. 1–23, doi:10.1006/jeth.1996.0035.
- [47] Stephen Morris (1997): *Alternative definitions of knowledge*. In Michael O. L. Bacharach, editor: *Epistemic logic and the theory of games and decisions*, Kluwer Academic Publishers, pp. 217–233, doi:10.1007/978-1-4613-1139-3\_10.
- [48] John von Neumann & Oskar Morgenstern (1944): *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.
- [49] Ariel Rubinstein & Asher Wolinsky (1990): *On the logic of “agreeing to disagree” type results*. *Journal of Economic Theory* 51(1), pp. 184–193, doi:10.1016/0022-0531(90)90057-Q.
- [50] Dov Samet (1990): *Ignoring ignorance and agreeing to disagree*. *Journal of Economic Theory* 52(1), pp. 190–207, doi:10.1016/0022-0531(90)90074-T.
- [51] Leonard J. Savage (1954): *The Foundations of Statistics*. John Wiley and Sons.
- [52] Burkhard C. Schipper (2013): *Awareness-dependent subjective expected utility*. *International Journal of Game Theory* 42(3), pp. 725–753, doi:10.1007/s00182-012-0321-2.
- [53] Burkhard C. Schipper (2014): *Preference-based unawareness*. *Mathematical Social Sciences* 70, pp. 34–41, doi:10.1016/j.mathsocsci.2013.04.003.
- [54] Burkhard C. Schipper (2014): *Unawareness—A gentle introduction to both the literature and the special issue*. *Mathematical Social Sciences* 70, pp. 1–9, doi:10.1016/j.mathsocsci.2014.03.002.
- [55] Burkhard C. Schipper (forthcoming): *Awareness*. In Hans van Ditmarsch, Joseph Y. Halpern, Wiebe van der Hoek & Barteld Kooi, editors: *Handbook of Logics for Knowledge and Belief*, London: College Publications, doi:10.2139/ssrn.2401352.
- [56] Dana Scott (1970): *Advice on Modal Logic*. In Karel Lambert, editor: *Philosophical Problems in Logic: Some Recent Developments*, Dordrecht: D. Reidel, pp. 143–173, doi:10.1007/978-94-010-3272-8\_7.
- [57] Krister Segerberg (1971): *An Essay in Classical Modal Logic*. *Filosofiska Studier* 13, Uppsala: Uppsala Universitet.

- [58] Giacomo Sillari (2006): *Models of Awareness*. In Giacomo Bonanno, Wiebe van der Hoek & Michael Wooldridge, editors: *Logic and the Foundations of Game and Decision Theory (LOFT 7)*, Amsterdam: Amsterdam University Press.
- [59] Giacomo Sillari (2008): *Quantified Logic of Awareness and Impossible Possible Worlds*. *The Review of Symbolic Logic* 1, pp. 514–529, doi:10.1017/S1755020308090072.

## A Material Supporting Section 2

### A.1 Stronger Assumptions about Knowledge

The model used in Theorem 2 already validates a number of attractive axioms on knowledge, suggesting that strengthening DLR’s constraints on knowledge is unlikely to yield an interesting triviality result. In particular, the following axioms are valid on the model:

Distribution:  $(Kp \wedge Kq) \rightarrow K(p \wedge q)$

Anti-Necessitation:  $\neg K \perp$

Reflexivity:  $Kp \rightarrow p$

Positive Introspection:  $Kp \rightarrow KKp$

We can show more systematically that any strengthening of the axioms of knowledge which rules out unawareness does so trivially in the same way as Negative Introspection does. On the very mild assumption that the agent doesn’t know the contradiction  $\perp$ , we can characterize the conditions under which a given model for the knowledge of an agent can be extended to an unawareness model in which the agent is unaware of a given  $p$  at a given point  $\alpha$  in which DLR’s three axioms are valid. Let a model  $M'$  extend a knowledge model  $M = \langle \Omega, k \rangle$  just in case  $M' = \langle \Omega, k, a \rangle$  for some function  $a : 2^\Omega \rightarrow 2^\Omega$ .

**Theorem 6.** *Let  $M = \langle \Omega, k \rangle$  be a knowledge model,  $\alpha \in \Omega$  and  $E \subseteq \Omega$  such that Anti-Necessitation is valid in  $\alpha$ .  $M$  has an extension such that Plausibility,  $KU$ -Introspection and  $AU$ -Introspection are valid in  $\alpha$  and  $\alpha \notin a(E)$  if and only if*

- (i)  $\alpha \in \neg k(E) \cap \neg k - k(E)$ , and
- (ii) there is an event  $F$  such that  $a \in F \cap \neg k(F) \cap \neg k - k(F)$ .<sup>10</sup>

*Proof.* Assume first that (i) and (ii). Let  $a : 2^\Omega \rightarrow 2^\Omega$  be defined so that  $a(E) = a(F) = \neg F$ , and  $a(G) = \Omega$  for all other events  $G$ , and consider the model  $M' = \langle \Omega, k, a \rangle$ . It is routine to verify that Plausibility,  $KU$ -Introspection and  $AU$ -Introspection are valid in  $\alpha$ , and  $\alpha \notin a(E)$ , appealing to Anti-Necessitation in the proof for  $KU$ -Introspection.

For the converse, note that (i) follows the validity of Plausibility in  $\alpha$ . For (ii), let  $F = \neg a(E)$ . Then  $\alpha \in F$ , by  $AU$ -Introspection,  $\alpha \notin a(F)$ , and so by Plausibility,  $\alpha \in \neg k(F) \cap \neg k - k(F)$ .  $\square$

In particular, as long as the constraints on knowledge allow for there to be an event  $E$  and a state  $\alpha \in E$  such that  $\alpha \in \neg k(E) \cap \neg k - k(E)$ , standard state space models and DLR’s three axioms will not preclude non-trivial unawareness.

<sup>10</sup>This result also holds if we replace Plausibility by  $n$ -Plausibility, for all natural numbers  $n$ , and (i) and (ii) by the correspondingly iterated conditions.

## A.2 Stronger Assumptions about Awareness

These results demonstrate that no plausible strengthening of the axioms governing knowledge will reinstate triviality. But what if we strengthen the axioms on awareness themselves?

To investigate this issue formally, extend the language  $L$  by a unary operator  $CK$  for common knowledge. To define its interpretation on a model  $M = \langle \Omega, k, a \rangle$ , derive the following functions on events:  $k^1(E) = k(E)$ ,  $k^{n+1}(E) = kk^n(E)$ , and  $ck(E) = \bigcap_{1 \leq n < \omega} k^n(E)$ .

$$\llbracket CK\varphi \rrbracket_{M,v} = ck(\llbracket \varphi \rrbracket_{M,v})$$

With this, we consider the following additional axioms on awareness:

$$CK\text{-Plausibility: } Ap \rightarrow CK(U p \rightarrow (\neg K p \wedge \neg K \neg K p))$$

$$CK\text{-}KU\text{-Introspection: } Ap \rightarrow CK(\neg KU p)$$

$$CK\text{-}AU\text{-Introspection: } Ap \rightarrow CK(U p \rightarrow U U p)$$

These additional axioms are also compatible with non-trivial unawareness. In fact, they are valid in state  $\alpha$  of the example in the proof of Theorem 2. More generally, Theorem 6 can be extended straightforwardly to these three additional axioms, given the weak assumption that Necessitation and Anti-Necessitation are valid:

**Theorem 7.** *Let  $M = \langle \Omega, k \rangle$  be a knowledge model in which Necessitation and Anti-Necessitation are valid,  $\alpha \in \Omega$  and  $E \subseteq \Omega$ .  $M$  has an extension such that Plausibility,  $KU$ -Introspection,  $AU$ -Introspection,  $CK$ -Plausibility,  $CK$ - $KU$ -Introspection and  $CK$ - $AU$ -Introspection are valid in  $\alpha$  and  $\alpha \notin a(E)$  if and only if*

$$(i) \alpha \in -k(E) \cap -k - k(E), \text{ and}$$

$$(ii) \text{ there is an event } F \text{ such that } \alpha \in F \cap -k(F) \cap -k - k(F).^{11}$$

*Proof.* We establish (i) and (ii) as in the proof of Theorem 6. Assuming (i) and (ii), we define  $a$  as in the proof of Theorem 6, where it is noted that Plausibility,  $KU$ -Introspection and  $AU$ -Introspection are valid in  $\alpha$ , and  $\alpha \notin a(E)$ . For the  $CK$ -conditions, consider any event  $G$  such that  $\alpha \in a(G)$ . Then by construction of  $a$ ,  $a(G) = \Omega$ . Therefore  $a(G) \cup H = \Omega$  for any event  $H$ , and so  $\alpha \in ck(a(G) \cup H)$  by Necessitation, which establishes the validity of  $CK$ -Plausibility and  $CK$ - $AU$ -Introspection in  $\alpha$ . For  $CK$ - $KU$ -Introspection, note that by Anti-Necessitation,  $k - a(G) = \emptyset$ , so  $-k - a(G) = \Omega$ , from which  $\alpha \in ck(-k - a(G))$  follows again by Necessitation.  $\square$

## B Proof of Theorem 4

*Proof.* Since the formulas derivable in this calculus form a classical model logic in the sense of [57], we can apply the standard canonical model construction technique; in particular, consider the smallest canonical model (see [7, chapter 9], especially p. 254). Consider any formula  $\varphi$  not provable in the above calculus, and let  $\Gamma$  be the set of subformulas of  $\varphi$  closed under Boolean combinations. A standard filtration of the canonical model through  $\Gamma$  produces a finite model in which  $\varphi$  is false. It is routine to prove that the neighborhood function for  $A_i$  associates with each state a field of sets; since the model is finite this field is generated by an equivalence relation, as required. The above filtration can be chosen in such a way as to preserve the transitivity of the relation for  $K_i$ ; reflexivity is preserved by any filtration (see, e.g. [7, chapter 3], especially p. 106, or [4] p. 80).

The above argument also establishes that the logic thus axiomatized has the finite model property and so is decidable.  $\square$

<sup>11</sup> Again, we can extend this result to  $n$ -Plausibility for all  $n$  analogously to the extension in the previous footnote.