Unawareness and Implicit Belief: Possible Worlds Models of Informational Content

Edward Elliott*

School of Philosophy, Religion and History of Science
University of Leeds

Abstract

Possible worlds models of belief have difficulties accounting for *unawareness*, the inability to entertain (and hence believe) certain propositions. Accommodating unawareness is important for adequately modelling epistemic states, and representing the informational content to which agents have access given their explicit beliefs. In this paper, I use neighbourhood structures to develop an original multi-agent model of explicit belief, awareness, and informational content, along with an associated sound and complete axiom system. I defend the model against the seminal impossibility result of Dekel et al. (1998), according to which three intuitive conditions preclude non-trivial unawareness on any 'standard' model of knowledge or belief. I argue that at least one of these conditions is implausible when applied to a model of belief. The plausibility of the other two rests on further questions regarding the scope and granularity of mental content. Finally, I show that, once we've jettisoned the least plausible of these conditions, it's possible to strengthen the remainder while retaining non-trivial unawareness within a possible worlds model of belief with unawareness.

Keywords: Awareness · Epistemic logic · Implicit belief · Possible worlds semantics · Neighbourhood structures

§1. Introduction

Sally is an agent, who may or may not be ideally rational. In her capacity as an agent, Sally has what we'll refer to broadly as *epistemic attitudes* towards a range of propositions. There will be some propositions that Sally knows, some she believes, and some that she's not sure about. Question: for *any* proposition φ, must Sally have some epistemic attitude or other towards φ?

Most will want to say that she need not. To know, believe, or be uncertain about a proposition requires the capacity to *entertain* that proposition, to represent it one way or another in thought—and there may be some propositions that Sally has never entertained, will never entertain, and perhaps *cannot* entertain.

For example, suppose that Sally belongs to an isolated tribe deep in the Amazon, which has had no contact with any other cultural group for hundreds

*Email: e.j.r.elliott@leeds.ac.uk. Draft of 29th January. Comments welcome.

of years. We wouldn't want to say that Sally knows, believes, or is uncertain about the proposition that we would express were we to utter the sentence 'The habitat of the eastern grey kangaroo extends as far north as the Cape York Peninsula'. It would furthermore be unnatural to say that Sally is agnostic or has suspended judgement on the matter of eastern grey kangaroo habitats. Suspension of judgement is usually understood as a mental act one performs upon entertaining some proposition, when one thinks that there are insufficient reasons to believe one way or the other. But Sally is not even *aware* that there's a content here for her to suspend judgement about. She may not have the resources necessary to think the thought, let alone decide whether she has sufficient evidence to believe it.

Let's admit, then, that there may be some propositions of which Sally is not aware. I use this phrase stipulatively: to be aware of a proposition φ is to have the basic representational resources required to entertain φ-thoughts or otherwise have what were above called epistemic attitudes towards φ.[1] One need not be presently and consciously attending to a proposition in order to be aware of it, and the entertaining itself might require significant mental effort—indeed, it may even require some degree of cognitive idealisation. So, for example, Sally may be aware of some massive conjunction of disjunctions of conditionals of various kinds between $\varphi, \psi, \neg\varphi$, and $\neg\psi$ merely by virtue of being able to entertain φ and ψ, even though entertaining such a complicated thought would push her working memory beyond its limits.

I am, therefore, using 'awareness' with a more determinate meaning than it's often used in the literature. For example, Fagin and Halpern (1988) are intentionally ambiguous with the term. For some of their discussion, it used in connection with the notion of *entertainability* that I've characterised here. In other cases, it marks a distinction between the beliefs a subject is consciously attending to and her 'background' beliefs; in still other cases to distinguish between that which can and cannot be derived from a base of stored beliefs within a given time. Given the kinds of conditions that awareness operators are standardly taken to satisfy (e.g., SYM, CON, PLA^a and the *AGPP* principle; see §2), and that awareness is often characterised in the first instance as "a lack of ability to conceive" (e.g., in Heifetz et al. 2008; Schipper 2014; Walker 2014), I suspect that many authors intend to use 'awareness' as I do. However, detailed exposition on the matter is rare and sometimes conflicting, making confident interpretation difficult. The model developed in §3 will be focused on capturing awareness as I've here characterised it, but the basic structures are flexible enough to accommodate other notions in the vicinity, such as *attentional* understandings of awareness.

It has long been recognised that the standard possible worlds models of belief and knowledge in the style of Hintikka (1962) do a poor job of accommodating unawareness.[2] Roughly, the issue is this. We begin with a rich space of possible

[1] It is worth emphasising the distinction in English between being aware/unaware *that* something is the case, and being aware/unaware *of* some thing. It is the latter we are interested in. The former is synonymous with believing a true proposition: Sally is aware *that* φ iff φ is true and Sally believes it; Sally is unaware that φ iff φ is true and Sally doesn't believe it. On the other hand, one is aware *of* a proposition in the way one is aware of an idea, an argument, or an oncoming train: by being able to entertain thoughts involving the object of one's awareness.

[2] See esp. (Fagin and Halpern 1988) for an early discussion of the issues. For a general introduction to key works in modelling unawareness, see (Schipper 2015).

worlds, Ω, subsets of which are taken to represent the contents of our attitudes.[3] Every agent Sally can then be associated with a binary relation R_s on Ω which for each world ω (effectively) picks out a proposition $R_s(\omega)$ containing all and only those worlds which Sally considers possible at ω. We then say that Sally *believes* that φ at ω iff φ is true at every world in $R_s(\omega)$. Problem: if $R_s(\omega)$ is even the least bit specific, then a vast number of things will be true at every world in $R_s(\omega)$, and we can't expect Sally to believe *all* of them—and not just because she may not be a very good deductive reasoner. Make Sally as logically gifted as you like: if she can't think certain thoughts then it's misleading to say that she believes them.

Consider the following very simple example. There are only two atomic propositions, φ and ψ, and exactly four possible worlds $\omega_1, \omega_2, \omega_3, \omega_4$ in Ω corresponding to the different combinations of φ and ψ:

ω_1 $\varphi \wedge \psi$	ω_2 $\varphi \wedge \neg\psi$
ω_3 $\neg\varphi \wedge \psi$	ω_4 $\neg\varphi \wedge \neg\psi$

Suppose that $R_s(\omega_1) = \{\omega_1, \omega_2\}$. This is just the set of worlds where φ is true, so we can say that Sally believes φ. Now presumably, if Sally believes φ then she can entertain φ; and if she can can entertain φ then she can distinguish between φ and $\neg\varphi$. So it looks fair to say that Sally is able to entertain $\neg\varphi$ at ω_1. Likewise, since she can entertain φ and $\neg\varphi$, she can entertain $\varphi \wedge \neg\varphi$ and $\varphi \vee \neg\varphi$. The latter is true at every world in $R_s(\omega_1)$, and it's reasonable enough to expect that Sally believes $\varphi \vee \neg\varphi$. At the very least, she can certainly reason her way to that conclusion given her other beliefs without much difficulty.

However, suppose that Sally is wholly unaware of the $\psi/\neg\psi$ distinction. Perhaps specifying the distinction requires some concept that Sally lacks. Would we be happy to say that Sally believes $\varphi \vee \psi$ and $\varphi \vee \neg\psi$ at ω_1? Both are true at every world in $R_s(\omega_1)$. But for Sally to believe $\varphi \vee \psi$ or $\varphi \vee \neg\psi$ would require her to make distinctions in thought which, *ex hypothesi*, she is unable to entertain.

Before we move on, two points of note are in order. First, you may get the intuition that it would be equally wrong to say that Sally is able to entertain $\psi \vee \neg\psi$, even though $\psi \vee \neg\psi$ picks out just the same set of worlds as $\varphi \vee \neg\varphi$ does. To that extent, you will probably think that possible worlds accounts of mental content suffer from problems relating to hyperintensionality more generally. And

[3] By my use of 'possible worlds,' I mean to exclude specifically worlds which are either not maximally specific, or inconsistent with classical logic. Much of what I say will not hang on any specific account of what possible worlds are; though see §5.3 for more discussion. For the sake of concreteness, the reader may wish to think of Ω as a set of maximal consistent (w.r.t. a consequence relation at least as strong as classical logic) sets of sentences of an appropriately rich formal language.

you may well might be right. (I will have more to say on this in §5.3-5.4.) But the problem I want to highlight here is separable from matters of granularity. For note that, even if you accept what I will call a *coarse-grained* account of content, such that $\psi \vee \neg\psi$ and $\varphi \vee \neg\varphi$ are by virtue of their logical equivalence one and the same object of thought just described in two different ways, then you will still need to deal with the possibility of unawareness. It's not plausible to say that Sally believes $\varphi \vee \psi$ under *any* description.

Second, you may think that the relational model has no real problems with unawareness at all—once it's interpreted in the right way. After all, it is frequently said that the model is best understood as a way of representing a subject's *implicit beliefs*, where this is understood to capture something like the informational content of the explicit beliefs the subject actually has stored somewhere in her head. On this interpretation, there need be no problem with claiming that Sally *implicitly* believes $\varphi \vee \psi$ and $\varphi \vee \neg\psi$, under the assumption that she *explicitly* believes φ: the former are built into the informational content implicit in the latter.

I want to be clear that I have no objections to interpreting the relational model in this way, or to this way of characterising what 'implicit beliefs' are. If that's the kind of thing you want to represent, then so be it. But I do want to introduce a distinction when thinking about the *informational content* of our epistemic attitudes, which underscores the need for incorporating some notion of awareness into our formal models. Say that in a *broad* sense, the informational content of a set of (explicit) attitudes consists in anything and everything that's entailed by the contents of those attitudes, whether separately or in conjunction. It is in this sense that it's unproblematic to say that Sally's belief that φ "contains the information" that $\varphi \vee \psi$ and $\varphi \vee \neg\psi$. After all, the information *is* there, just waiting for someone with the appropriate representational resources to pull it out. But Sally is *not* one of those with the required resources. For her, some of that broad informational content is inaccessible. It is not available for Sally to use in reasoning, inference and decision making. In a perfectly good sense the information is *invisible*: Sally may not even know that there's a distinction she's not considered, and no amount of reasoning with the distinctions she does have will lead her to recognise what she's been missing. And information which is wholly invisible to Sally is, from her perspective, no information at all.

So, it looks like there is room for a second, *narrow* way of understanding informational content, one that's specifically dependent upon awareness. Arguably, it's the latter kind of narrow informational content that will be most useful in understanding how Sally navigates the world on the basis of the information stored in her explicit beliefs. Note that under this narrow understanding of informational content, there is room for Stalnaker's (1991) useful distinction between content which is more or less readily accessible for cognitive application. It's plausible to think that the way we store information about the world makes a difference to how easily we might access it on a given occasion, and consequently that on any given occasion we probably only attend to a fragment of the total information we've got represented across the full range of our epistemic attitudes. As Stalnaker notes, recognising the limits of attention and access goes some of the way towards solving the problem of logical omniscience (cf. also Egan 2008; Elga and Rayo 2015). But any such a solution cannot be complete without also recognising the role that awareness has as a precondition for informational access *simpliciter*. Once we've formulated an adequate model for representing the distinction between contents we do and do not have access to in principle, we can get to work on modelling varying degrees of access as a result of various processing limitations.

What's needed at this stage, then, is a model which appropriately distinguishes on the one hand between Sally's explicit epistemic attitudes, and the narrow informational content contained within those attitudes, on the other. In this paper, I will focus primarily on the attitude of belief, and to a lesser extent, knowledge. In §2-3, I characterise the class of what we will call *Possible-Worlds Awareness* (or PWA) models, along with an associated axiom system for reasoning about them within a simple modal language. As the name suggests, PWA models retain the coarse-grained approach to mental content traditionally associated with possible worlds semantics, and are consequently subject to the influential triviality theorem of (Dekel et al. 1998). In §4, I outline a lightly modified version of Dekel et al.'s theorem, and in §5, I critically discuss the intuitions motivating it. As we will see, the force of their result depends much on how we interpret the model, and (moreover) on questions regarding the nature and scope of mental content. Finally, in §6, I show how a PWA model of belief can be used to capture strengthened versions of most of Dekel et al.'s conditions (and all of the most plausible ones).

§2. Preliminaries

To construct our model, we will first need a formal language. Given a finite set of agents $\mathbf{Ag} = \{1, \ldots, n\}$ and a countable set of atomic propositions Φ (with typical element p), we characterise $\mathcal{L}^{AXI}(\Phi)$ by the following grammar:

$$\varphi \ := \ p \mid \neg\varphi \mid \varphi \wedge \psi \mid A_i\varphi \mid X_i\varphi \mid I_i\varphi,$$

where $p \in \Phi$ and $i \in \mathbf{Ag}$. We abbreviate with '$\vee$,' '$\rightarrow$,' and '$\leftrightarrow$' in the usual ways, and we will let '$\top$' stand for '$p \vee \neg p$', and '$\bot$' for '$p \wedge \neg p$.'

'$A_i\varphi$' should be read as saying that 'i is aware of φ'. I will leave the interpretation of the X operators ambiguous for now, between an explicit *knowledge* and an explicit *belief* reading. Likewise, the I operators can be used to stand for either implicit knowledge or implicit belief, depending on how we choose to read the Xs. (I will have more to say on these two interpretations as we go along.) The intended notion of implicit knowledge/belief that we want here is tied to the narrow informational content of the subject's explicit attitudes.

As noted, the usual way of generating a semantics for epistemic modals involves assigning a relation R_i to each agent i which picks out those worlds i considers possible at each given world ω. However, for added flexibility in characterising agents' awareness and moreover for distinguishing between explicit and implicit beliefs, I will use the strictly more general neighbourhood structures of (Scott 1970) and (Montague 1970). A *neighbourhood model* M consists in a set of worlds Ω; a set of *neighbourhood functions* $\mathcal{N}_i$ (one for each $i \in \mathbf{Ag}$) which map each world ω in Ω to a (potentially empty) set of (potentially empty) subsets of Ω; and a propositional valuation function $\pi : \Phi \mapsto 2^{\Omega}$. The satisfaction conditions for $\varphi \in \mathcal{L}^{X}(\Phi)$ can then be given as follows:

$M, \omega \models p$ iff $\omega \in \pi(p)$, for $p \in \Phi$
$M, \omega \models \neg\varphi$ iff it's not the case that $M, \omega \models \varphi$
$M, \omega \models \varphi \wedge \psi$ iff $M, \omega \models \varphi$ and $M, \omega \models \psi$
$M, \omega \models X_i\varphi$ iff $\{\omega' : M, \omega' \models \varphi\} \in \mathcal{N}_i(\omega)$

If we let '$\|\varphi\|^M$' (the *truth set of φ in M*) refer to the set of worlds ω such that $M, \omega \models \varphi$, then the final clause can be stated a little more perspicuously as:

$M, \omega \models X_i\varphi$ iff $\|\varphi\|^M \in \mathcal{N}_i(\omega)$

Thus, we can use neighbourhood functions to characterise directly those propositions which we want to say i knows/believes at a given world ω. Note that there is no inbuilt assumption, for any φ and any ω, that $\|\varphi\|^M$ must belong to $\mathcal{N}_i(\omega)$; i.e., that $X_i\varphi$ must be true at any world. Likewise, we don't assume that if $\|\varphi\|^M \in \mathcal{N}_i$ and ψ is true at every world in $\|\varphi\|^M$, then $\|\psi\|^M \in \mathcal{N}_i$. Explicit beliefs and knowledge need not be closed under implication. Indeed, the logic associated with neighbourhood models is very weak, consisting of just:

PROP All classical propositional tautologies
MP From φ and $\varphi \to \psi$, infer ψ
REP From $\varphi \leftrightarrow \varphi'$, infer $\psi \leftrightarrow \psi[\varphi/\varphi']$

Where '$\psi[\varphi/\varphi']$' denotes any sentence that results from the replacement of zero or more instances of φ in ψ with φ'. (For example, $(X_i(p \vee \neg p))[p/q]$ can refer to $X_i(p \vee \neg p)$, $X_i(q \vee \neg p)$, $X_i(p \vee \neg q)$, or $X_i(q \vee \neg q)$.) Following Chellas (1980), we will refer to any axiom system which contains PROP, MP, and REP as *classical*.

The flexibility afforded by neighbourhood structures will also be useful in characterising awareness, which we will represent by an additional set of functions $\mathcal{A}_i$, which we will call *awareness functions*. Like neighbourhood functions, every awareness function is a mapping from worlds to sets of sets of worlds. However, in keeping with the intended interpretation of the model, we will want to place a few additional constraints upon each agent's awareness function and its relationship with neighbourhood functions.

First of all, and most obviously, an agent's state of awareness cannot float free of her explicit beliefs. By definition, if an agent isn't aware of a proposition, then she cannot explicitly believe it or know it. For reasons that will become apparent below, we'll refer to this as PLA^a:

PLA^a $X_i\varphi \to A_i\varphi$

In Definition 1 (below), we capture PLA^a with the condition (*pla*), that $\mathcal{N}_i(\omega) \subseteq \mathcal{A}_i(\omega)$, for all ω.

Furthermore, on any plausible account of awareness as propositional entertainability, we will need to take into consideration the productivity and systematicity of thought. We can safely assume that the capacity to think φ-thoughts goes hand in hand with the capacity to think $\neg\varphi$-thoughts. This should be uncontroversial, being common ground for most accounts of propositional mental representation. Likewise, and for similar reasons, anyone who can think φ-thoughts and ψ-thoughts can, in principle at least, also think $(\varphi \wedge \psi)$-thoughts. Hence, SYM $A_i\varphi \to A_i\neg\varphi$

CON $(A_i\varphi \wedge A_i\psi) \to A_i(\varphi \wedge \psi)$

We ensure that these axioms hold by assuming, respectively, that $\mathcal{A}_i$ is closed under complements *(sym)* and binary intersections *(con)*. So, for all ω, $\mathcal{A}_i(\omega)$ is either empty or a countably additive Boolean algebra.

Note that we will *not* be assuming that an awareness of $\varphi \wedge \psi$ implies an awareness of φ and awareness of ψ. In any classical logic, this would quickly lead to triviality: given SYM and CON, if i is aware of any φ, then she's aware of $\varphi \wedge \neg\varphi$, and hence (by REP) aware of $\psi \wedge \neg\psi$ for arbitrary ψ. Awareness of any φ would imply awareness of every ψ, which is clearly unacceptable. In this respect, PWA models are quite unlike many other models of unawareness in the literature, where it's very common to suppose that an agent is aware of $\varphi \wedge \psi$ only if she's aware of φ and ψ (see, e.g., Halpern 2001; Heifetz et al. 2006; Li 2009; Schipper 2015; Cozic 2016).

More generally, it is very common for models of unawareness to satisfy the property of *awareness generated by primitive ropositions* (*AGPP*), i.e., that i is aware of every primitive proposition in $\Psi \subseteq \Phi$ if and only if she is aware of every $\varphi \in \mathcal{L}^{AXI}(\Psi)$. *AGPP* is obviously not going to play nicely with any remotely coarse-grained account of content, for essentially the reasons just outlined. But the *AGPP* principle seems deeply implausible—in fact, it's *especially* implausible under a fine-grained account of content. For i to be aware of A_jp, for instance, she must be aware not only of the primitive proposition p, but also the agent j and the attitude of *awareness.*[4] Sally's capacity to think φ-thoughts does not come with an automatic capacity to think that Bob is aware of, or believes that, φ. So, while there may be some uses for a model wherein every agent is aware of every other agent and every variety of propositional attitude under examination, *AGPP* is clearly too strong as a *general* principle for modelling awareness. Thus, we will want to reject each of the following consequences of *AGPP*:

DIS $A_i(\varphi \wedge \psi) \to (A_i\varphi \wedge A_i\psi)$

AIA $A_i\varphi \to A_iA_j\varphi$

AIX $A_i\varphi \to A_iX_j\varphi$

AII $A_i\varphi \to A_iI_j\varphi$

By contrast, SYM and CON (in conjunction with PLA^a) are only strong enough to get us the result that if an agent i is aware of every φ in a set $\mathcal{X} \subseteq \mathcal{L}^{AXI}(\Phi)$, then i will also be aware of everything in the closure of $\mathcal{X}$ under $\neg$ and $\wedge$. And this seems much more reasonable: if i is worthy of the title of 'agent' and has the representational resources to entertain every primitive proposition in Ψ, then she *ipso facto* has the resources to entertain logical compounds of those propositions.[5]

[4] Proviso: I am assuming that A_jp is epistemically contingent. See §5.3 for discussion.

[5] It's worth noting that by removing *(pla)*, *(sym)* and/or *(con)* from Definition 1 below, we can drop PLA^a, SYM and/or CON. This would be useful if we wanted to capture other notions of 'awareness' that are sometimes discussed in the literature for which these conditions may seem inappropriate. As noted above, 'awareness' is sometimes understood in terms of *attention*: the contents an agent is 'aware of' are just those beliefs she is attending to at the time. On this interpretation, we would want to assume that $A_i\varphi$ implies $X_i\varphi$ (and not vice versa), and there is no strong reason to think that CON (and perhaps SYM) is valid for attentional awareness.

Finally, and primarily for simplicity, I will make one further assumption about the relationship between A_i and X_i, expressed by the following:

N_{AX} $\qquad A_i\top \to X_i\top$

In the context of the other conditions, if i is aware of anything at all, then she is aware of $\top$. In context, then, N_{AX} states that i has awareness only if she explicitly knows or believes $\top$. Where the condition REP has already been accepted, this is quite weak. It is hard to imagine an agent worthy of the title who does not accept even the simplest of tautologies, and the most popular theories of coarse-grained mental content imply that agents don't even have a choice about whether to believe $\top$. We capture this in the model by *(nax)*, that if Ω belongs to $\mathcal{A}_i$, then it belongs to $\mathcal{N}_i$. Building *(nax)* into the definition of a PWA model simplifies the relationship between explicit and implicit attitudes by ruling out the possibility of awareness without any explicit attitudes, without going so far as to commit us to the more common (and much stronger) axiom N_X (discussed in §4).[6]

In §6, I will discuss the addition of further constraints on awareness functions, but now let us move on to the representation of narrow informational content. It is here that the awareness function does most of its work. If i believes each of a collection of propositions $\varphi_1, \ldots, \varphi_n$ which jointly imply ψ, then we will say that i implicitly knows/believes ψ whenever she's aware of ψ. We model this by first characterising (for each world ω) the largest set of worlds $P \subseteq \Omega$ consistent with all of i's explicit attitudes at ω as the intersection of all the sets of worlds in $\mathcal{N}_i(\omega)$. We then say that i implicitly knows/believes any proposition Q such that $P \subseteq Q$ just in case she is aware of Q. So, Sally implicitly believes φ just in case φ is something Sally can entertain that's implied by the conjunction of her explicit beliefs. Thus, the awareness function $\mathcal{A}_i$ acts as a filter or sieve over the propositions which the agent implicitly knows/believes in the broad sense.

§3. Coarse-Grained Awareness Structures

With that out of the way, the class of PWA models can be defined:

Definition 1 A model $M = (\Omega, \{\mathcal{N}_i\}_{i\in\mathbf{Ag}}, \{\mathcal{A}_i\}_{i\in\mathbf{Ag}}, \pi)$ belongs to the class of PWA models $\boldsymbol{\mathcal{M}}$ iff:

(1) Ω is a non-empty set

(2) $\mathcal{N}_i$ and $\mathcal{A}_i$ are functions from Ω to 2^{2^Ω}, satisfying (for all $\omega \in \Omega$):

(pla) $\quad P \in \mathcal{N}_i(\omega)$ only if $P \in \mathcal{A}_i(\omega)$
(sym) $\quad P \in \mathcal{A}_i(\omega)$ only if $\Omega \setminus P \in \mathcal{A}_i(\omega)$
(con) $\quad P_1 \in \mathcal{A}_i(\omega)$ and $P_2 \in \mathcal{A}_i(\omega)$ only if $P_1 \cap P_2 \in \mathcal{A}_i(\omega)$
(nax) $\quad \Omega \in \mathcal{A}_i(\omega)$ only if $\Omega \in \mathcal{N}_i(\omega)$

(3) π is a propositional valuation function

Definition 2 then characterises what it is for an element of $\mathcal{L}^{AXI}(\Phi)$ to be true at a given world ω in a PWA model M (i.e., $M, \omega \models \varphi$):

[6] If the reader is unwilling to accept N_{AX}, we can do without it by (i) removing condition *(nax)* from Definition 1, and (ii) replacing the axiom N_{AX} in Σ with the weaker axiom $A_i\top \to I_i\top$. The soundness and completeness proofs will be left almost completely unchanged.

Definition 2 For all $\varphi \in \mathcal{L}^{AXI}(\Phi)$,

$M, \omega \models p$ iff $\omega \in \pi(p)$, for $p \in \Phi$
$M, \omega \models \neg\varphi$ iff it's not the case that $M, \omega \models \varphi$
$M, \omega \models \varphi \wedge \psi$ iff $M, \omega \models \varphi$ and $M, \omega \models \psi$
$M, \omega \models X_i\varphi$ iff $\|\varphi\|^M \in \mathcal{N}_i(\omega)$
$M, \omega \models A_i\varphi$ iff $\|\varphi\|^M \in \mathcal{A}_i(\omega)$
$M, \omega \models I_i\varphi$ iff $\|\varphi\|^M \in \mathcal{A}_i(\omega)$ and $\bigcap \mathcal{N}_i(\omega) \subseteq \|\varphi\|^M$

Furthermore, we say that φ is *valid in* M iff $M, \omega \models \varphi$ for every $\omega \in M$; and *valid in* $\boldsymbol{\mathcal{M}}$ (i.e., $\boldsymbol{\mathcal{M}} \models \varphi$) iff φ is valid in every $M \in \boldsymbol{\mathcal{M}}$.

The classical logic associated with the class of PWA models can then be axiomatised by the following system, which we'll label Σ:

PROP	All classical propositional tautologies
SYM	$A_i\varphi \to A_i\neg\varphi$
CON	$(A_i\varphi \wedge A_i\psi) \to A_i(\varphi \wedge \psi)$
XI	$X_i\varphi \to I_i\varphi$
IA	$I_i\varphi \to A_i\varphi$
K$_{AI}$	$(I_i\varphi \wedge I_i(\varphi \to \psi)) \to (A_i\psi \to I_i\psi)$
N$_{AX}$	$A_i\top \to X_i\top$
MP	From φ and $\varphi \to \psi$, infer ψ
REP	From $\varphi \leftrightarrow \varphi'$, infer $\psi \leftrightarrow \psi[\varphi/\varphi']$

Say that φ is a *theorem of* Σ (i.e., $\vdash_\Sigma \varphi$) just in case φ is either an axiom of Σ or can be derived from the axioms by finite applications of the inference rules. We can then show that φ is a theorem of Σ if and only if it is valid in the class of PWA models:

Theorem 1 *Σ is sound and complete with respect to $\boldsymbol{\mathcal{M}}$ and $\mathcal{L}^{AXI}(\Phi)$; i.e., for all $\varphi \in \mathcal{L}^{AXI}(\Phi)$,*

$$\boldsymbol{\mathcal{M}} \models \varphi \text{ iff } \vdash_\Sigma \varphi$$

Proof. See Appendix A. □

Σ as developed so far is best understood as a logic for belief rather than knowledge, as it imposes no requirement of veridicality on X (and I). Where a minimal conception of knowledge takes it to be a species of veridical (or 'non-delusional,' 'factive') belief, we would want our logic to include:

T$_I$ $\quad I_i\varphi \to \varphi$

Which, in combination with XI, will get us:

T$_X$ $\quad X_i\varphi \to \varphi$

We can ensure T$_I$ and T$_X$ by adding a reflexivity condition of the form:

(t) $\quad$ If $P \in \mathcal{N}_i(\omega)$, then $\omega \in P$

Since, under (t), $X_i\varphi$ is true only at worlds where φ is true, and $I_i\varphi$ is only true at worlds ω where $A_i\varphi$ and $\bigcap \mathcal{N}_i(\omega) \subseteq \|\varphi\|^M$, the addition of ($t$) will also get us that $I_i\varphi$ is true at ω only if φ is too (as $\omega \in \bigcap \mathcal{N}_i(\omega)$).

Considered primarily as a way of representing the relationship between implicit and explicit belief (i.e., ignoring the awareness operator), PWA models resemble the *local reasoning structures* of Fagin and Halpern (1988, pp. 58ff), intended to represent a 'fragmented' belief system of the kind discussed by Lewis (1982). Indeed, PWA models are essentially generalised local reasoning structures with the addition of awareness functions acting as filters over the supersets of $\bigcap \mathcal{N}_i(\omega)$. (Fagin and Halpern define '$M, \omega \models X_i\varphi$' in a manner that's stronger than is standard for neighbourhood structures, leading to a stronger logic.)

PWA models are also similar to the more recent model of implicit and explicit belief developed by Velázquez-Quesada (2013). Velázquez-Quesada's construction works by taking a neighbourhood function $\mathcal{N}_i$ defined for a finite space Ω (interpreted as specifying i's explicit beliefs), and using it to systematically construct another function $\mathcal{N}_i^\star$ which contains Ω and is closed under supersets and binary intersections, interpreted as specifying i's implicit beliefs. In outline, this construction of $\mathcal{N}_i^\star$ is quite similar to the construction of $\|I_i\varphi\|^M$ as a subset of $\{\omega : \bigcap \mathcal{N}_i(\omega) \subseteq \|\varphi\|^M\}$, though there are some important differences. One of these concerns the axiom XI, which is not valid on Velázquez-Quesada's class of models. XI is important where the I operator is understood to represent the informational content contained in an agent's explicit beliefs—every explicit belief that φ obviously contains the information that φ.

In terms of the relationship between implicit belief and awareness, the use of awareness functions in PWA models is conceptually very similar to the way such functions are used in the *Kripke structures for general awareness* of Fagin and Halpern (1988)—though (famously) the latter models use syntactically-valued awareness functions to generate extremely fine-grained distinctions between states of awareness. And finally, in the presence of (*sym*) and (*con*), and with (*pla*) dropped and stronger conditions placed on the $\mathcal{N}_i$, PWA models also have much in common with the *partitional models* of Fritz and Lederman (2015), which are intended to capture a strong logic of belief and its interaction with awareness within the context of a coarse-grained approach to content.

§4. On the Impossibility of Unawareness

In this section, I outline a lightly modified version of Dekel et al.'s well-known impossibility result. It centres on three straightforward assumptions about the relationship between unawareness and the epistemic attitudes:

PLA $\quad \neg A_i\varphi \rightarrow (\neg X_i\varphi \wedge \neg X_i\neg X_i\varphi)$

AUR $\quad A_i\neg A_i\varphi \rightarrow A_i\varphi$

XUI $\quad \neg X_i\neg A_i\varphi$

Dekel et al. themselves describe PLA (for 'Plausibility') as the most plausible of their three axioms (hence the name), and many in the literature have followed them in treating it as fundamental to an adequate understanding of awareness. See, for instance, (Modica and Rustichini 1994, 1999), (Heifetz et al. 2006), (Chen et al. 2012), (Walker 2014), (Cozic 2016), though cf. (Halpern 2001), where PLA is derived as a condition on unawareness only under certain assumptions.

Given a classical logic, PLA is of course just the conjunction of PLA^a with:

PLA^b	$\neg A_i\varphi \rightarrow \neg X_i \neg X_i \varphi$

PLA^b says that if i is unaware of φ, then she cannot believe that she doesn't believe φ. AUR (for 'AU Reflection') says that if i is capable of thinking, of herself, that she is unaware of some proposition φ, then she is *ipso facto* capable of thinking about φ directly. The intuition behind PLA^b and AUR is the same, and is easiest to grasp if we view awareness and belief as relations that hold between a thinking subject and those propositions she entertains—the idea being that anyone who can think that *a is not R-related to b* has the conceptual resources to think in terms of a, R, and b. Thus, for PLA^b: if i can believe that she's not belief-related to φ, the she can think in terms of φ directly. Likewise, for AUR: if i can entertain the idea that she is not awareness-related to φ, then she can entertain φ directly.

Finally, XUI (for 'XU Introspection') says it's impossible for i to believe that she is unaware of some specific proposition φ. To be clear: i might be aware that there *exists* some propositions she's not aware of, and which some other agent may be aware of. All of the conditions we discuss in this paper are consistent with saying this much. But i cannot be aware of a *specific* state of unawareness that she has, towards a particular proposition φ. To express the existentially quantified thoughts, we would need more than the non-quantificational language that I'm employing here.[7]

In §5.1-5.3, we will examine the plausibility of these conditions in more detail; for now, let us focus on the impossibility result:

Theorem 2 *Suppose Σ^* is a classical logic which includes* PLA, AUR, *and* XUI. *Then, if Σ^* includes* N$_X$, *then for every φ,*

$$\vdash_{\Sigma^*} A_i\varphi,$$

and if Σ^* includes MON, *then for all φ and ψ,*

$$\vdash_{\Sigma^*} \neg A_i\varphi \rightarrow \neg X_i\psi$$

Where:

N$_X$	$X_i\top$
MON	From $\varphi \rightarrow \psi$, infer $X_i\varphi \rightarrow X_i\psi$

Proof. From AUR, PLA^b: $\neg A_i\varphi \rightarrow \neg A_i\neg A_i\varphi$ and $\neg A_i\neg A_i\varphi \rightarrow \neg X_i\neg X_i\neg A_i\varphi$. With XUI and REP, this gets us $\neg A_i\varphi \rightarrow \neg X_i\top$. So, $\neg A_i\varphi$ is inconsistent with N$_X$. Furthermore, MON implies that for any ψ, $X_i\psi \rightarrow X_i\top$; or equivalently, $\neg X_i\top \rightarrow \neg X_i\psi$. So, $\neg A_i\varphi \rightarrow \neg X_i\psi$. □

Thus, for any classical logic Σ^*, the very presence of a non-trivial A_i operator is incompatible with PLA, AUR, XUI, and at least one of N$_X$ or MON. Moreover, neither N$_X$ nor MON (nor PLA^a, for that matter) are essential to generating a problem, as the following corollary indicates:

[7] For relevant work in this regard, see (Halpern and Rêgo 2009), (Sillari 2008), (Walker 2014). Given the straightforward nature of PWA models, I would not anticipate any difficulties in extending them to allow for quantification over propositional variables. I have not done this since the issues relevant to the present paper arise already with non-quantificational languages.

Corollary 1 *Suppose Σ^* is a classical logic which includes* PLA^b*,* AUR*, and* XUI*. Then, for all φ, $\vdash_{\Sigma^*} \neg A_i\varphi \to \neg X_i\top$.*

And this is already a very troubling result—unawareness shouldn't preclude belief in simple propositional tautologies!

We can generalise the badness a little further, by noting that XUI can be broken down into two separate conditions:

XUI^a $\quad \neg A_i\varphi \to \neg X_i \neg A_i\varphi$
XUI^b $\quad A_i\varphi \to \neg X_i \neg A_i\varphi$

XUI^a is a consequence of AUR and PLA^a, so what XUI actually brings to Theorem 2 is better understood through XUI^b. Hence,

Corollary 2 *Suppose Σ^* is a classical logic which includes* PLA^a*,* PLA^b*,* AUR*, and* XUI^b*. Then, for all φ, $\vdash_{\Sigma^*} \neg A_i\varphi \to \neg X_i\top$.*

What's worse, if we add SYM, CON, and N$_{AX}$ back into the mix, then we get that if i is unaware of anything, then she's unaware of everything:

Corollary 3 *Suppose Σ^* is a classical logic which includes* PLA^a*,* PLA^b*,* AUR*,* XUI^b*,* SYM*,* CON*, and* N$_{AX}$*. Then, for all φ and ψ, $\vdash_{\Sigma^*} \neg A_i\varphi \to \neg A_i\psi$.*

Proof. From Corollary 2, $\neg A_i\varphi \to \neg X_i\top$, and from N$_{AX}$, $\neg X_i\top \to \neg A_i\top$. We've seen that in any classical logic with PLA^a, SYM, and CON, $A_i\varphi$ implies $A_i(\varphi \vee \neg\varphi)$, and hence $A_i\top$. So, $\neg A_i\top \to \neg A_i\psi$; and for all φ and ψ, $\neg A_i\varphi \to \neg A_i\psi$. □

§5. An Analysis of PLA, AUR, and XUI

For the sake of the arguments to follow, suppose that we're happy to stick with a classical logic of the epistemic attitudes. (I will defend this supposition in §5.4.) Since PLA^a is surely secure, the only question left for anyone modelling awareness is which of PLA^b, AUR, or XUI^b we ought to give up. In the following sections, I will argue that the decision should rest partly on how we want to interpret X, on what kind of 'coarse-grained' account of content we adopt, and on whether there exist any unthinkable propositions. We begin with a look at XUI^b under the different interpretations of X.

§5.1 XUI and Non-Veridical Belief

Suppose first of all that we want X to represent knowledge, or indeed any other veridical epistemic attitude. (For example, it's plausible that *rational certainty* is veridical: correctly updating on one's evidence should never lead one to become immutably convinced of a falsehood.) Since XUI^b follows immediately from T$_X$, and T$_X$ just is the veridicality condition, there is no plausible way out of the triviality result via XUI^b under this kind of interpretation.

Suppose on the other hand that we want X to be interpreted as belief.[8] Under this interpretation, XUI^b says that if i is aware of φ, then she does not explicitly

[8] It's worth noting that, although Dekel et al. express their result only in terms of 'knowledge,' they don't presuppose that T$_X$ is valid in general, and (more importantly) the result has been widely taken to apply to all epistemic models which use Aumann structures or something closely analogous, including those intended to represent non-veridical states. See, e.g., (Sillari 2008, p. 516), (Schipper 2013), (Cozic 2016, p. 3) and the models outlined in (Schipper 2015).

believe that she's unaware of φ. This does not seem especially plausible. It would be perhaps *odd* for anyone to have false beliefs about their own state of awareness. After all, if you're aware of φ, then you *arguably* have access to all the evidence you need to know that you're aware that φ. XUI^b, however, says that an agent *cannot* falsely believe she's in some state of unawareness. And non-ideal agents can have all sorts of absurd and unjustified beliefs that fly in the face of their evidence, and beliefs about awareness ought to be no different.

> **Example 1** In an attempt to naturalise her ontology of mind, Sally ends up committed to an error theory about folk psychology. According to this theory, all talk of being 'aware of,' 'entertaining,' or even 'believing' abstract entities like propositions is bunk, a manner of thinking associated with an outdated and generally false folk psychology.[9] Sally has thought a lot about folk psychology and about her stance on it, so she's aware of the proposition *Sally does not believe that folk psychology is true*. However, she doesn't believe that she's aware of that proposition—in fact, she's certain that she isn't.

Letting φ the proposition that *Sally does not believe that folk psychology is true*, we have in this situation $A_s\varphi \wedge X_s\neg A_s\varphi$, and a counterexample to XUI^b.

So here is our first lesson: while XUI^b is unavoidable if our goal is to model a veridical attitude like knowledge, it does not look like we should want to keep it around if our goal is to model non-veridical belief. And note that the point here is independent of how fine- or coarse-grained we want mental content to be: XUI^b is false for belief under any position on the granularity of mental content. Furthermore, in §6, I will prove that supplementing Σ with strengthened versions of PLA^b and AUR is consistent with non-trivial awareness in the absence of XUI^b. So, where our goal is to develop a logic of a non-veridical epistemic attitude with unawareness, we have a way out of the triviality result.

§5.2 The Plausibility of 'Plausibility'

Suppose, therefore, that our goal is to model knowledge. In this subsection, I will argue that any fan of coarse-grained content ought to reject PLA^b.

Either AUR holds or it does not. Suppose first that it does. Then, for some φ, suppose that $X_i\neg A_i\varphi$. Knowledge presupposes awareness, so if $X_i\neg A_i\varphi$, then $A_i\neg A_i\varphi$, which then implies $A_i\varphi$. However, from T$_X$, we know that $\neg A_i\varphi$. Contradiction. Assuming AUR establishes the existence of non-contingent knowledge states: for any i and any φ, it's impossible for i to know $\neg A_i\varphi$. Now with this established, PLA^b should look less plausible than it might have on first appearances. If $\neg X_i\neg A_i\varphi$ is necessary, then $X_i\neg X_i\neg A_i\varphi$ is just $X_i\top$, and we have $X_i\top$ at any world where the agent is aware of anything. PLA^b then implies that $X_i\neg X_i\neg A_i\varphi \rightarrow A_i\neg A_i\varphi$, where the consequent by AUR implies $A_i\varphi$. But there are numerous ways an agent can come to know the necessary proposition $\top$, many of which won't require an awareness of φ for arbitrary φ. Of course,

[9] An error theorist about a domain of discourse holds that at least all positive, first-order, atomic and non-trivial or non-analytic sentences in the domain are meaningful (truth-apt), yet systematically false. An error theorist about folk psychology would deny the truth of any sentence of the form $A_i\varphi$, $X_i\varphi$, or $I_i\varphi$, but may accept the truth of, e.g., $\neg A_i\varphi$, $A_i\varphi \rightarrow A_i\varphi$, or $A_i\varphi \vee \psi$. In $\mathcal{L}^{AXI}(\Phi)$, $A_i\varphi$, $X_i\varphi$, or $I_i\varphi$ are not 'atomic,' but each corresponds to an atomic sentence in the ordinary languages where folk psychological discourse usually occurs.

the foregoing is just to restate the triviality result, but one person's *ponens* is another's *tollens*. If AUR holds, then PLA^b looks deeply implausible on any coarse-grained account of content, precisely because on that kind of account we need to make allowances for the fact that propositions like $\top$ and $\bot$ can be known under a myriad of guises.

Suppose now that AUR is invalid. Assuming SYM, this can only be the case if there exists some ψ such that ψ and $A_i\varphi$ are equivalent, yet $A_i\psi$ does not imply $A_i\varphi$. Suppose that such a ψ exists. This is just to suppose that there are ways to entertain $A_i\varphi$ under a guise that doesn't require awareness of φ. (In the same way, for example, that one can entertain $p \vee \neg p$ under a guise that doesn't require awareness of p, by entertaining $q \vee \neg q$.) Now, if it's possible for $A_i\psi \wedge \neg A_i\varphi$ to be true, then (since $A_i\psi \leftrightarrow A_i\neg\psi$ and $\neg A_i\varphi \leftrightarrow \neg\psi$) it's possible for $A_i\neg\psi \wedge \neg\psi$ to be true; and, *presumably*, in that case it's also possible for $X_i\neg\psi \wedge \neg\psi$ to be true. That is, any counterexample to AUR can be expected to generate a counterexample to PLA^b. (We will see this evidenced below, where we consider potential counterexamples to AUR.) This should be unsurprising, since AUR and PLA^b rest on precisely the same basic intuitions about awareness of propositional attitude relations.

The only way the above argument would fail would be if all circumstances under which $A_i\neg\psi \wedge \neg\psi$ is true are such as to preclude the possibility of $X_i\neg\psi$. These would have to be very strange circumstances indeed. Since $\neg A_i\neg\psi$ already implies $\neg X_i\neg\psi$, this would mean that $\neg\psi \to \neg X_i\neg\psi$—which together with the fact that $X_i\neg\psi \to \neg\psi$ implies that $X_i\neg\psi$ is impossible. So suppose that $X_i\neg\psi$ is impossible. Then by PLA^b, $\neg A_i\neg\psi$ implies $\neg X_i\neg X_i\neg\psi$, the former of which is equivalent to $\neg A_i\psi$ and the latter of which is equivalent to $\neg X_i\top$. Given N$_{AX}$, $\neg X_i\top \to \neg A_i\top$, hence $\neg A_i\psi \to \neg A_i\top$. And finally, given $A_i\psi \to A_i\top$, we reach the conclusion that $A_i\psi \leftrightarrow A_i\top$. But we know that ψ is not equivalent to $\top$, for that would make $\neg A_i\varphi$ impossible, which by hypothesis it is not. And we know that ψ is not equivalent to $\bot$, for that would imply that $\neg X_i\top$ is necessary, which is clearly absurd. Hence, for PLA^b to avoid falling foul of the whatever counterexamples exist for AUR, we would need that for every φ such that $A_i\neg A_i\varphi \to A_i\varphi$ is *in*valid, and the following are also valid:

1. $A_iA_i\varphi \leftrightarrow A_i\top$
2. $X_iA_i\varphi \to \bot$

I think it is safe to assume that no such φ exists. It's implausible that there are any propositions of the form $A_i\varphi$ the awareness of which is implied by the awareness of anything whatsoever, which *also* cannot be known even in the (possible) circumstances where it's true. One can readily imagine propositions φ such that $A_iA_i\varphi$ *might* arguably be true whenever $A_i\top$ is true, such as $\varphi = \top$; but any such example is of a proposition that i can surely know she's aware of whenever she's aware of it. Moreover, under N$_{AX}$, the proposition $A_iA_i\varphi$ is just the proposition *i is aware of something* (it's true whenever $\mathcal{A}_i$ is non-empty), and this is certainly something that i can know.

So here is our second lesson: given a coarse-grained account of content, PLA^b turns out to be quite implausible in the context of T$_X$. Either AUR is valid and so generates counterexamples to PLA^b, or it's invalid, in which case any of its counterexamples are equally counterexamples to PLA^b. In either case, PLA^b

should be dropped. The only remaining issue for the coarse-grained content theorist is whether she ought to also reject AUR.

§5.3 AUR and Two Conceptions of Content

Because AUR does not interact in any interesting way with T_X, we can consider its plausibility independently of how we interpret the X operators. Nevertheless, AUR turns out to be the trickiest of Dekel et al.'s three conditions to deal with, and doing so will involve opening more than one can of worms. The number of issues here necessitates some brevity; I do not pretend to have stated the last word on any of them. Ultimately, I will argue that counterexamples to AUR are potentially available on one conception of coarse-grained content, while on another (and possibly more useful for present purposes) conception, AUR is arguably valid.

5.3.1 Metaphysical Conceptions of Content

Let's begin with what we can call the *metaphysical* conception of coarse-grained content, according to which if $\varphi \leftrightarrow \psi$ holds as a matter of metaphysical necessity, then φ and ψ denote one and the same content. Ω on this picture can be thought of as the space of metaphysical possibilities, such that to be aware of (or believe that) *Water is H_2O* is just to be aware of (or believe that) *H_2O is H_2O*. Most readers, when they think of a coarse-grained account of content, will likely have this kind of conception in mind; it is the kind of account most typically (but not necessarily) associated with externalism about content. Below, I'll discuss how AUR fares on an alternative *epistemic* conception of the kind more familiar to internalist and two-dimensionalist approaches to mental content, but for now we focus on the metaphysical conception.[10]

On the metaphysical conception, counterexamples to AUR will take the form of a proposition ψ such that $\neg A_i\varphi \leftrightarrow \psi$ is metaphysically necessary, but $A_i\psi \rightarrow A_i\varphi$ is not. On first appearances, counterexamples of this form should be easy to find: for a given proposition ψ, there will often be many ways for different agents to entertain something metaphysically equivalent to ψ despite having very different conceptual resources. I suspect that the relevant counterexamples do exist, but they are not *easy* to establish.

To start with the obvious suggestion, there may be propositions φ which are *necessarily* unentertainable. If any such proposition exists, then $\neg A_i\varphi$ is necessary, and there will be (many) ways to entertain $\neg A_i\varphi$ without entertaining φ. But are there any such φ? It's certainly not obvious that there are.[11] Let me briefly consider two arguments for the existence of unthinkable propositions.

On the first (and easily most common) kind of argument, one might think that some propositions cannot be entertained by us as a result of our cognitive limitations. For instance, due to limited resources of memory and processing,

[10] Thanks to Harvey Lederman for highlighting to me the importance of discussing AUR under the metaphysical conception, and for discussion on these issues more generally. I owe examples 3 and 4 below to him.

[11] For a recent and thorough treatment of several arguments relating to the present discussion, see (Hofweber 2016). Note that Hofweber's discussion is centred on the issue of whether there are any aspects of reality which cannot be expressed in language or thought by beings *like us*, given the way we are. His concern is not whether there exist propositions that are necessarily unentertainable for a given agent.

there may be some φ which are simply too complex or specific for an ordinary agent like Sally to entertain. This example, I think, clearly fails: these are contingent limitations to human representational capacities, and there are no reasons to suspect that there are no metaphysically possible worlds where they have been overcome (including, potentially, in our own future). But there are considerations in the vicinity which deserve more weight. In particular, one might think that there are hard-wired limitations on the kinds of phenomenological experiences we might undergo (cf. Nagel 1986, pp. 90ff). Given the kind of being I am, perhaps I cannot—as a matter of metaphysical necessity—know what it is like to be a bat. If it is essential to Sally that she is a human, and it's essential to being human that one is "wired up" in *this* way, such as to preclude the possibility of her undergoing certain kinds of experiences, then we might have a counterexample to AUR. This paper is not an exercise in metaphysics, so I will leave these questions hanging.

The second kind of argument seeks to show that there aren't enough thinkable propositions to go around. Perhaps the best known of these comes from Lewis (1986, pp. 104-7). If we are the right kind of functionalists about contentful mental states, then there can be no more thinkable thoughts than there are possible functional roles. However, Lewis asserts, there are probably no more than countably many functional roles that could be used to define our mental states, while there are probably at least $\beth^3$ sets of possible worlds. So, it turns out, there are unthinkable propositions—in fact, most propositions are unthinkable.

I won't try to tackle this argument head-on. Set aside what you think about the premises needed to get the cardinality argument going, and note that the conclusion, if true, wouldn't immediately generate trouble for AUR. There are at most countably many φ in $\mathcal{L}^{AXI}(\Phi)$, and mere concerns about cardinality don't provide a reason to think that any of *these* are necessarily unthinkable. (There are many propositions which have no correlate in $\mathcal{L}^{AXI}(\Phi)$; that these may or may not be unentertainable is no problem for AUR.) Fix the language such that for each p in Φ, there's a world where i can entertain p. Then, it's plausible that i can (in some world or other) entertain any φ in $\mathcal{L}^{AXI}(\Phi)$. After all, $\mathcal{L}^{AXI}(\Phi)$ consists in simple Boolean combinations of primitive propositions and the closure of those under a small number of simple propositional attitude operators. Cardinality arguments do not give us a reason to think that we cannot find a functional role for every φ in $\mathcal{L}^{AXI}(\Phi)$.

So it's not obvious that there are any necessarily unentertainable propositions. (The discussion of the next case gives further reasons for doubt.) Can we instead find a contingent proposition ψ such that $\psi \leftrightarrow A_i\varphi$ is necessary, but $A_i\psi$ does not necessitate $A_i\varphi$?[12] Consider the following case:

> **Example 2** Sally has just made first contact with an alien from Gzorp, with whom she's attempting to communicate. In it's own language, the Gzorpian asserts that *Sally is not aware of the proposition that skirnobs are poisonous*, where a *skirnob* is a variety of Gzorpian fruit. Sally does not understand, of course, but she knows that the Gzorpian just asserted something, and that whatever it is, it's probably true.

[12] Counterexamples to AUR can exist only for where ψ is contingent or impossible: if $A_i\varphi$ is necessary, then $A_i\neg A_i\varphi \to A_i\varphi$ is trivial.

Let φ denote the proposition *skirnobs are poisonous*, and let ψ pick out that set of worlds where *the proposition the Gzorpian actually just expressed is true*. Then it's presumably the case that $A_s\psi$, and $\psi \leftrightarrow \neg A_s\varphi$. If it's also the case that $\neg A_s\varphi$, then we have a counterexample to AUR.

For the case to work, it needs to be true that Sally has no way of entertaining the thought that *skirnobs are poisonous* under any mode of presentation whatsoever. So, for instance, we must suppose that she doesn't have the representational resources to think anything from the following (obviously non-exhaustive) list:

(a) The fruit the Gzorpian just referred to is poisonous
(b) The fruit with purple stripes and pink feathers is poisonous
(c) The fruit I would be thinking about, were I in this brain state at that time, is poisonous
(d) The fruit, an instance of which is exactly 65.36325 light years away from me bearing 92.5473° at an elevation of 30.1323°, is poisonous
(e) The things in the category my community refers to when they make the sound /fru:t/, an instance of which is exactly 65.36325 light years away from me bearing 92.5473° at an elevation of 30.1323°, are poisonous
(f) The things referred to by the sound /skə:nɒb/ in the language generally spoken on the planet 65.36325 light years away from me bearing 92.5473° at an elevation of 30.1323° are poisonous

Give her enough sortal concepts and a way to represent arbitrary relative distances, directions, and times, and there won't be many objects that actually exist/existed/will exist, or properties that are/were/will be instantiated, that Sally won't be able to think about under some mode of presentation. She never *will* have thoughts under such modes of presentation, but she *could*—and that's all that's required for awareness. Perhaps Sally could be aware of the proposition *Sally is not aware of the proposition that skirnobs are poisonous* without being aware of *skirnobs are poisonous*, but she would have to be quite representationally impoverished indeed.

The problem, of course, is that the flexibility in constructing alternative guises for entertaining $A_i\varphi$ goes hand-in-hand with a flexibility in constructing alternative guises for entertaining φ. Awareness can come *very* cheap on the metaphysical conception. It's easy to imagine ways of picking out the set of worlds where $A_i\varphi$ holds which don't specifically mention φ; it's a lot less easy to imagine the capacity to entertain the former without *any* capacity to entertain something equivalent to the latter. One more example:

> **Example 3** As it turns out, the state of *being aware of the proposition that skirnobs are poisonous* can be necessarily identified with a particular psychophysical state, S. Sally is a brilliant neuroscientist, and can represent herself as being in state S.

This time, letting ψ denote that set of worlds where Sally is in state S, $\psi \leftrightarrow A_s\varphi$ is necessary. And since $A_s\psi$, so we also have $A_s\neg\psi$.

But do we have $A_s\varphi$? This would be easier to argue for an internalist about content, as then S could be identified with some internal (presumably neurological) state, and it's more plausible that Sally could represent an arbitrary neurological state without being able to represent that *skirnobs are poisonous*.

Merely looking at a brain in that state and thinking *my brain is like that* should suffice. But internalism sits poorly with a metaphysical conception of content, and few adherents to the latter would want to commit themselves to the former. Externalists will have a much tougher time of it: on that kind of picture, S would presumably be a highly disjunctive psychophysical state, where at least some of the disjuncts would involve causal connections between the agent and skirnobs (or something that belongs to a causal chain that starts with skirnobs). Being able to pick out *that* kind of state without being able to pick out *skirnobs* in some way or another is not obviously possible.

There are additional cases we could consider, involving, e.g., Burgean social externalism and cases of entertaining a thought without being an expert in the relevant concepts merely by being situated in the relevant linguistic community, brute metaphysical necessities, and so on. I will not try to consider them all. What is clear is that there is a case—or rather, several cases—to be made for the existence of counterexamples to AUR on a metaphysical conception of content.

5.3.2 Epistemic Conceptions of Content

So much for the metaphysical conception. Let's now consider an alternative, *epistemic* conception of coarse-grained content.[13] Say that φ is *epistemically possible* just in case it cannot be ruled out a priori. The thought that φ can then be said to correspond to an epistemic possibility, a way the world might be for all one might know a priori. Given this, let Ω designate the space of maximally specific epistemic possibilities; i.e., for each ω and every φ, ω either a priori entails φ or $\neg\varphi$, and ω entails both φ and ψ only if $\varphi \wedge \psi$ is epistemically possible. Then, φ and ψ are *a priori equivalent* just in case φ and ψ are entailed by all the same ω. All logical truths and falsehoods will be a priori equivalent, as will be, e.g., *bachelors are bachelors* and *bachelors are unmarried available men*. And, most importantly, if two thoughts φ and ψ under two different guises are metaphysically equivalent, yet the equivalence of those guises is not a priori, then φ and ψ will correspond to distinct subsets of Ω.

Most theorists who make use of the epistemic conception are pluralists about content: they accept that the metaphysical conception is explanatorily useful for many purposes, but also that it cannot play all of the explanatory roles for which we might want a notion of content to play. The epistemic conception, in particular, seems better equipped to account for the phenomenon of cognitive significance. That *H_2O is H_2O* is trivial and easily discovered upon a priori reflection, but no amount of reasoning absent empirical evidence will get us to *water is H_2O*. Where our task is to model the kind of information an agent has available to her via reasoning from her explicit beliefs, the epistemic conception thus seems particularly apt. But lest it be said that the epistemic conception does not "really" give us a coarse-grained approach to content, let me note some points before moving on.

First, one should not assume that the epistemic conception of coarse-grained content involves merely supplementing the original space of metaphysically possible worlds with a number of metaphysically impossible worlds so as to let us

[13] I include here only a bare-bones development of the epistemic conception. More detailed developments can be found in (Jackson 1998, 2009) and (Chalmers 2002, 2011, 2006). I have defended the epistemic conception elsewhere; see (Elliott et al. 2013).

distinguish between, e.g., those worlds where *Hesperus is Hesperus* and those where *Hesperus is Phosphorus*. For one thing, at least some metaphysical possibilities seem to be a priori false. For instance, where 'watery' designates the property of

> *being the clear, potable liquid around here that fills the lakes and oceans and falls from the sky as rain*,

then *water is watery* is arguably a priori, but it's certainly not metaphysically necessary that H_2O *is watery*. Or a less controversial example: *if something is watery, then the stuff that is actually watery is watery* is clearly a priori, but it is not metaphysically necessary. The epistemic conception of content is not merely the metaphysical conception with a few extra distinctions between worlds. Indeed, on so-called 'one-spaceist' views (e.g., Jackson 2009, and Chalmers 2006, p. 82) the set of (centred) metaphysically possible worlds and the set of maximally specific epistemic possibilities are one and the same: every (centred) world can be thought of in one of two ways: as an hypothesis about how the world might be for all one might know a priori, or as an hypothesis about how it could have been given the way things actually are. According to one-spaceism, then, there's no sense in which the epistemic conception of content is more (or less) fine-grained than the metaphysical conception—contents on either conception are simply subsets of a single space of worlds, Ω.

Supposing then that classical logic is a priori, if we adopt the epistemic conception we will end up with a model on which PROP, MP, and REP are valid, but which also lets us draw some distinctions between contents under different modes of presentation that are unavailable on the metaphysical conception. For instance, on the epistemic conception, no pair from the list (a)-(f) earlier of guises for thinking that *skirnobs are poisonous* are a priori equivalent, so each will hold relative to a different subset of Ω. Moreover, cases like Example 2 and Example 3 fail to generate counterexamples to AUR, for the proposition ψ that Sally is supposed to be aware of in those cases is not a priori equivalent to $A_i\varphi$. To consider just the latter example, it is certainly not a priori that *Sally is in brain state S* if and only if *Sally is aware of the proposition that Sally is aware of the proposition that skirnobs are poisonous*. Awareness is not as cheap on the epistemic conception. It should be clear that other purported counterexamples which rely on *a posteriori* metaphysical identities or rigidified definite descriptions will fail on the epistemic conception for similar reasons.[14]

Moreover, there is a general reason for thinking that counterexamples won't arise once we've got a notion of content that cuts as fine as cognitive significance—for how could Sally entertain some content ψ that has the very same cognitive significance as *Sally is awareness-related to* φ, without having the conceptual resources to represent *Sally*, *awareness*, and φ? Supposing that ψ is epistemically contingent, then AUR looks essentially right. Unlike $\top$, there are only so many ways to entertain the thought that Sally is aware of φ under a mode of

[14] The same points apply to cases that involve linguistic deference and social externalism, though I have not discussed these. When a non-expert thinks to themselves, *I have arthritis in my thigh*, their grasp of *arthritis* is distinct from the understanding of an expert. Roughly, it is something like *the disease the experts refer to when they say 'arthritis'*. What the non-expert knows a priori when they know something "merely by being situated in a linguistic community" is quite different than what the experts know. Cf. Chalmers 2002, §9.

presentation that's a priori equivalent to *Sally is aware of* φ, and it's reasonable to expect that they all go hand-in-hand with the capacity to entertain φ itself.

Furthermore, there's a stronger case to be made that there are no φ such that it's a priori impossible that Sally is aware of φ. Arguments from cognitive limitations don't seem to get any grip: for any limitations that are supposedly essential Sally *qua* human being, it is not *a priori* for Sally that she is subject to those limitations. (For all she knows a priori, she *could* have been a bat.) And the same points that apply to the argument from cardinality apply here.

Before we move on, let me consider one final case for the existence of a proposition that's a priori unentertainable:

> **Example 4** Sally decides to let 'Silly' name that proposition φ such that, from amongst those propositions of which she is unaware, is expressible by the shortest English sentence. She thinks to herself: *I am not aware of Silly.*

Assuming that Sally's designation succeeds (e.g., there is a unique φ that satisfies the description), then Sally's thought is certainly a priori. But we have to be careful here. The case does not establish the existence of a proposition φ such that it's a priori for Sally that she is not aware of φ, because it is not a priori what proposition 'Silly' picks out (if in fact it picks out anything at all). What Sally knows a priori is that if 'Silly' designates something, then whatever proposition it designates, she is not aware of that proposition. At one epistemic possibility ω_1, 'Silly' might designate φ_1; at ω_2, φ_2. Unless it's a priori what proposition 'Silly' picks out, there's no specific φ that Sally knows a priori she's not aware of, and AUR is safe.

§5.4 In Defence of Classical Logics

Obviously, a lot of what I've argued in §5.2 and §5.3 hangs on the acceptance of a coarse-grained account of content of some form or another. I don't expect to have convinced anyone who thinks that the contents of thought cut finer than (classical) logical equivalence, and to such a reader it may seem like I am seeing counterexamples to PLA^b and AUR where I really ought to be seeing counterexamples to REP. But that is itself an important third lesson of the present discussion: to the extent that incorporating awareness into classical logics of belief and knowledge seems to present a problem at all, it doesn't add any *specific* issues over and above the much more general concerns about hyperintensionality.

For one who's already accepted coarse-grained contents, it's not a problem to be told that there's something counterintuitive to saying that i can be aware of $\neg X_i \neg A_i \varphi$ even while unaware of φ, in the special circumstance where $\neg X_i \neg A_i \varphi$ is impossible. This is on a par with being told that one can know that $\varphi \vee \neg\varphi$, even while one is unaware of φ, and no coarse-grained content theorist is going to give up their position because of that!

Proponents of coarse-grained contents already have a suite of tools to help explain away the intuitions in favour of more fine-grained accounts of content. Given a Two-Dimensionalist approach to mental content and an appropriately characterised space of epistemically possible worlds, many problem cases for possible worlds semantics—e.g., the need to distinguish between metaphysically equivalent but epistemically non-equivalent water-beliefs and H_2O-beliefs—can be dealt with very naturally without deviating from the basic idea that contents are sets of possible worlds. Conversational pragmatics can help deal with linguistic intuitions about the substitutability of that-clauses within the context of propositional attitude verbs (Stalnaker 1978), and the 'fragmentation' of belief states (Stalnaker 1984; Lewis 1982, 1986; Fagin and Halpern 1988) helps deal with concerns relating to information access and logical closure properties.[15]

The biggest concern for classical logics of belief arise when they are taken in conjunction with a plausible theory of action; *viz.*, that agents will typically act so as to maximise their desire satisfaction given the way they believe the world to be (cf. Stalnaker 1991). If theories of coarse-grained content have empirically false behavioural implications then they must be rejected, and on the face of it the typical subject doesn't seem to act as we might expect given beliefs in every necessary truth and a typical set of desires. Thus the coarse-grained theorist has to tell us a story about why, for example, a mathematician might spend her days trying to work out whether or not the Riemann hypothesis is true. (She already knows the answer to *that* question; she just don't know whether the sentence 'The Riemann zeta function has its zeros only at the negative even integers and complex numbers with real part $\frac{1}{2}$' expresses a necessary truth—but then it's not so obvious what set of fragmented beliefs and desires we can plausibly attribute to rationalise the time spent discovering yet another complicated way to say $\top$.)

For AUR and PLA^b to generate similar kinds of empirical issues for the coarse-grained theory, we would need to have a case of an ordinary agent i with presumably ordinary desires who, with respect to some φ such that it's impossible for i to be aware of φ (or such that i cannot know that she is unaware of φ), tend to behave as if they don't believe that they're unaware of φ (or as if they don't believe that they don't know they're unaware of φ). But in what way does the typical agent generally fail to act so as to suggest that she doesn't have the relevant beliefs?

Examples are not easy to find, since (under REP) i will also believe that $\top \rightarrow \neg A_i\varphi$ (or $\top \rightarrow \neg X_i \neg A_i\varphi$), which will tend to mute any behavioural consequences that we might otherwise have expected to be generated by the stated beliefs. For instance, perhaps i doesn't like being unaware of anything; hence, for i, $\neg A_i\varphi$ represents an undesirable state of affairs. But if she knows that $\top \rightarrow \neg A_i\varphi$, then she won't try to *do* anything to improve her overall state of awareness in this respect. More generally, according to view of rational action that's supposed to generate the problem, actions are a response to the things agents believe they can change so as improve their situation. Taken in combination with a coarse-grained account of content, we shouldn't expect a belief in a necessary proposition to have interesting behavioural consequences: agents who believe a necessary proposition also believe that it's necessary, that it will be a fact of the world regardless of what they choose. So it's certainly not obvious that there would be any behavioural consequences of saying that i believes $\neg A_i\varphi$ which aren't in fact borne out by i's behaviour.

It may turn out that there is really is no way to assign plausible coarse-grained beliefs and desires so as to rationalise the actions of ordinary agents. If

[15] This final point is less pressing issue for PWA models given that explicit beliefs are not closed under implication, and implicit beliefs are only closed under implication relative to awareness.

so, we'll need to either adopt a more fine-grained approach to content, or revise our theory of action. I take this as an open empirical question. The success of standard models of decision-making—which generally make use of coarse-grained content, and which adhere more or less to the basic expected utility maximisation paradigm—provides a limited reason to think that the behavioural data can be accommodated within a classical logic. Or, at the least, they provide reasons to continue *modelling* contents using sets of possible worlds. Whatever issues such models may or may not have, they are independent of considerations arising from unawareness.

§6. Higher-Order Multi-Agent Awareness

It's one thing to deny the premises of an argument, and quite another to deny its conclusion. For all I've said, there may be many other routes to triviality than that taken by Theorem 2 and its corollaries. In this section, therefore, I want to expand upon the lessons of §5 and see how far PWA models can go towards recapturing the intuitions behind PLA^b and AUR within a model of non-veridical belief.

I assume that there are no non-contingent (implicit or explicit) belief or awareness states—and more generally that there are no counterexamples to PLA^b and AUR along the lines of those discussed above. Indeed, under this assumption I think we can go beyond those two conditions by adding the following three axioms to Σ:

ARA $\quad A_i A_j \varphi \rightarrow A_i \varphi$
ARX $\quad A_i X_j \varphi \rightarrow A_i \varphi$
ARI $\quad A_i I_j \varphi \rightarrow A_i \varphi$

Given SYM, ARA immediately implies AUR; and given SYM and PLA^a, ARX gets us to PLA^b. The three axioms in combination say that for all agents i and j, if i is aware of j's having some attitude regarding φ, then i is herself aware of φ. For the reasons discussed in §2, I think it would be a mistake to assume the converse of any of these three axioms.

Let Σ^+ refer to the system $\Sigma \cup \{\text{ARA}, \text{ARX}, \text{ARI}\}$, and let $\boldsymbol{\mathcal{M}}^+$ refer to that class of PWA models which satisfies the following additional constraints:

(ara) $\quad$ If $\{\omega' : P \in \mathcal{A}_j(\omega')\} \in \mathcal{A}_i(\omega)$, then $P \in \mathcal{A}_i(\omega)$
(arx) $\quad$ If $\{\omega' : P \in \mathcal{N}_j(\omega')\} \in \mathcal{A}_i(\omega)$, then $P \in \mathcal{A}_i(\omega)$
(ari) $\quad$ If $\{\omega' : P \in \mathcal{A}_j(\omega')$ and $\bigcap N_i(\omega) \subseteq P\} \in \mathcal{A}_i(\omega)$, then $P \in \mathcal{A}_i(\omega)$

The following is then easy to prove:

Theorem 3 Σ^+ *is sound and complete with respect to* $\boldsymbol{\mathcal{M}}^+$ *and* $\mathcal{L}^{AXI}(\Phi)$

Proof. See Appendix B $\square$

Moreover, there are models belonging to $\boldsymbol{\mathcal{M}}^+$ which allow for non-trivial awareness. To demonstrate this, we'll model a standard test case of speculative trade (as discussed in, e.g., Schipper 2015):[16]

[16] The reader may note that the reading of 'awareness' as *entertainability* in the following example is somewhat strained, and the example fits somewhat better with a reading of 'aware-

> **Example 5** Sally is the owner of a firm, and Bob is a potential buyer. Sally is aware of a potential lawsuit arising from an obscure legal technicality that would reduce the value of the firm significantly, though she is uncertain whether the lawsuit will occur. Bob is unaware of the lawsuit and Sally recognises this. Bob is aware of a potential innovation which would increase the value of the firm significantly, though he is uncertain whether the innovation will occur. Sally is unaware of the innovation and Bob recognises this.

For the model $M^e = (\Omega, \mathcal{N}_s, \mathcal{N}_b, \mathcal{A}_s, \mathcal{A}_b, \pi)$, let $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}$, and $\Phi = \{p, q\}$, with p representing the situation of the lawsuit and q representing the situation of the innovation. We will designate subsets P of Ω using the worlds by which they're constituted; hence, let $P_{n\dots m} = \{\omega_n, \dots, \omega_m\}$. We will suppose that $\pi(p) = P_{135}$ and $\pi(q) = P_{125}$. Finally, we let Sally's awareness and neighbourhood functions be defined as follows:

$$\begin{array}{ll}
\mathcal{A}_s(\omega_1) = \{\varnothing, P_5, P_{13}, P_{24}, P_{135}, P_{245}, P_{1234}, \Omega\} & \mathcal{N}_s(\omega_1) = \{P_{13}, \Omega\} \\
\mathcal{A}_s(\omega_2) = \varnothing & \mathcal{N}_s(\omega_2) = \varnothing \\
\mathcal{A}_s(\omega_3) = 2^\Omega & \mathcal{N}_s(\omega_3) = 2^\Omega \\
\mathcal{A}_s(\omega_4) = \{\varnothing, P_{34}, P_{125}, \Omega\} & \mathcal{N}_s(\omega_4) = \{\Omega\} \\
\mathcal{A}_s(\omega_5) = \{\varnothing, P_{24}, P_{135}, \Omega\} & \mathcal{N}_s(\omega_5) = \{\Omega\}
\end{array}$$

And Bob's functions:

$$\begin{array}{ll}
\mathcal{A}_b(\omega_1) = \{\varnothing, P_5, P_{12}, P_{34}, P_{125}, P_{345}, P_{1234}, \Omega\} & \mathcal{N}_b(\omega_1) = \{P_{12}, \Omega\} \\
\mathcal{A}_b(\omega_2) = 2^\Omega & \mathcal{N}_b(\omega_2) = 2^\Omega \\
\mathcal{A}_b(\omega_3) = \varnothing & \mathcal{N}_b(\omega_3) = \varnothing \\
\mathcal{A}_b(\omega_4) = \{\varnothing, P_{24}, P_{135}, \Omega\} & \mathcal{N}_b(\omega_4) = \{\Omega\} \\
\mathcal{A}_b(\omega_5) = \{\varnothing, P_{34}, P_{125}, \Omega\} & \mathcal{N}_b(\omega_5) = \{\Omega\}
\end{array}$$

The world of interest is ω_1, where we have a situation like the one described in the speculative trade example. Sally is aware of p but unaware of q, while Bob is aware of q but unaware of p. Sally believes that Bob isn't aware of p, while Bob believes that Sally isn't aware of q. Furthermore, Sally believes neither p nor $\neg p$, and Bob believes neither q nor $\neg q$. Most importantly, both Sally and Bob have non-trivial unawareness at ω_1 (and at ω_4, ω_5).

It's easy to check that the model satisfies *(pla)*, *(sym)*, *(con)*, and *(nax)*. It is less easy (and a little tedious) to check that M^e satisfies *(ara)*, *(arx)*, and *(ari)*, but those three conditions also hold. That M^e satisfies *(ara)* can be seen by noting first of all that:

$$\{\omega' : P \in \mathcal{A}_s(\omega')\} = \begin{cases} P_{1345}, & \text{if } P \in \{\varnothing, \Omega\} \\ P_{135}, & \text{if } P \in \{P_{24}, P_{135}\} \\ P_{13}, & \text{if } P \in \{P_5, P_{13}, P_{245}, P_{1234}\} \\ P_{34}, & \text{if } P \in \{P_{34}, P_{125}\} \\ P_3, & \text{otherwise} \end{cases}$$

ness' in terms of *attention.* As noted in §1, the interpretation of 'awareness' is often difficult, and one of the reasons for this difficulty is a mismatch between authors' characterisations of 'awareness' as given in the abstract and the kinds of examples in which it's used. I use the speculative trade case only because it is a standard test case for models of awareness in the literature—the details don't matter for the point I want to make. Sympathy is extended to those already dealing with the terminological confusions present in the literature.

$$\{\omega' : P \in \mathcal{A}_b(\omega')\} = \begin{cases} P_{1245}, & \text{if } P \in \{\varnothing, \Omega\} \\ P_{125}, & \text{if } P \in \{P_{34}, P_{125}\} \\ P_{12}, & \text{if } P \in \{P_5, P_{12}, P_{345}, P_{1234}\} \\ P_{24}, & \text{if } P \in \{P_{24}, P_{135}\} \\ P_2, & \text{otherwise} \end{cases}$$

That is, there are exactly five propositions $P \subseteq \Omega$ such that $P = \|A_s\varphi\|^{M^e}$ for some φ, and five P such that $P = \|A_b\varphi\|^{M^e}$ for some φ.

Focus first on $\mathcal{A}_s$. It's immediate that P_2, P_3, P_{12}, P_{1245}, and P_{1345} belong to $\mathcal{A}_s(\omega')$ only in the trivial case of ω_3, where $\mathcal{A}_s(\omega_3) = 2^\Omega$. This case obviously meets the requirements of (*ara*). Likewise, $P_{13} = \|A_s\varphi\|^{M^e}$ exactly when $\|\varphi\|^{M^e}$ equals P_5, P_{245}, P_{1234}, or P_{13}, and Sally is only aware of the first three propositions in the trivial case of ω_3, and (obviously) $P_{13} \in \mathcal{A}_s(\omega')$ for any world ω' such that $P_{13} \in \mathcal{A}_s(\omega')$. Next, $P_{135} = \|A_s\varphi\|^{M^e}$ whenever $\|\varphi\|^{M^e}$ equals P_{24} or P_{135}, whereas $P_{24} = \|A_b\varphi\|^{M^e}$ exactly when $\|\varphi\|^{M^e}$ equals P_{24} or P_{135}—and, by virtue of (*sym*), in every world where Sally is aware of P_{135} she's aware of P_{24} and *vice versa*. Similarly, $P_{34} = \|A_s\varphi\|^{M^e}$ when $\|\varphi\|^{M^e}$ equals P_{34} or P_{125}, while $P_{125} = \|A_b\varphi\|^{M^e}$ exactly when $\|\varphi\|^{M^e}$ equals P_{34} or P_{125}. Once again, (*sym*) guarantees that whenever Sally is aware of P_{34} she's aware of P_{125}, and *vice versa*. Thus, $\mathcal{A}_s$ meets the requirements of (*ara*).

A parallel argument for $\mathcal{A}_b$ then establishes that M^e satisfies (*ara*), and similar reasoning can be used to show straightforwardly that (*arx*) and (*ari*) are likewise satisfied.

From Example 5, Theorem 4 follows:

Theorem 4 *For all φ and ψ, $\boldsymbol{\mathcal{M}}^+ \not\models \neg A_i\varphi \to \neg X_i\top$ and $\boldsymbol{\mathcal{M}}^+ \not\models \neg A_i\psi \to \neg A_i\varphi$*

That is, even under the strengthened conditions (*arx*), (*ara*) and (*ari*), there are models in $\boldsymbol{\mathcal{M}}^+$ which can accommodate the possibility of non-trivial unawareness, in the sense that:

1. An agent can be unaware of φ and still believe $\top$
2. An agent can be unaware of φ without being unaware of everything

§7. Conclusion

The very large majority of epistemic or doxastic logics that have been developed over the past two decades which feature an awareness operator in some form or another have been non-classical. All of these logics are incompatible with a coarse-grained (or possible worlds) approach to content. This sets much the work on unawareness somewhat at odds with research elsewhere in formal epistemology, where possible worlds models of propositional content are still very much standard—and for good reason. Moreover, the present state of affairs leaves the advocate of possible worlds contents without an appropriate way to understand the impact that unawareness has on belief and informational content.

I have shown that it's possible to retain PLA and AUR—widely considered to be of special importance to any model of unawareness—within a possible worlds model of belief with unawareness. We can do this as long as we give up XUI, which I've argued we should do regardless of how fine-grained we take mental content to be. There are no *formal* reasons arising from PLA and AUR for not adopting a model that makes use of coarse-grained, sets-of-possible-worlds contents.

I have also argued that when it comes to modelling knowledge, the arguments in favour of PLA add nothing over and above already existing arguments against coarse-grained contents. Theorem 2 and its corollaries give us no reason to think that a classical logic with unawareness is any worse off than the classical logic of belief was already with respect to the puzzles associated with hyperintensionality. There is a rich body of philosophical work dealing with exactly these puzzles within the framework of traditional possible worlds semantics, and all the reason in the world to think that the very same work can be marshalled in support of a possible worlds model of awareness.[17]

Appendix A

The soundness part of Theorem 1 is straightforward by induction and left to the reader. To prove completeness we will construct a canonical PWA model.

For any $\varphi \in \mathcal{L}^{AXI}(\Phi)$, say that φ is consistent (relative to the system Σ) iff it's not the case that $\vdash_{\Sigma} \neg\varphi$; a finite set $\Gamma = \{\varphi_1, \dots, \varphi_n\}$ is consistent iff $\varphi_1 \wedge \cdots \wedge \varphi_n$ is consistent; and an arbitrary set $\Gamma = \{\varphi_1, \varphi_2, \dots\}$ is consistent iff every finite subset of Γ is consistent. Finally, say that Γ is a maximal consistent set (i.e., $\textsc{Max}\Gamma$) just in case Γ is consistent and maximal, in the sense that any strict superset of Γ is inconsistent.

On the ordinary way of constructing canonical models, the set of 'worlds' is just the set of all maximal consistent sets of formulas. For the present proof, however, it will be easier to include two 'worlds' for every maximally consistent set Γ. In this, I am applying a modified version of a strategy from Fagin and Halpern (1988). For the sequel, then, let $\Omega^0 = \{\Gamma^n : \textsc{Max}\Gamma,\ n = 0\}$, and $\Omega^1 = \{\Gamma^n : \textsc{Max}\Gamma,\ n = 1\}$, with $\Omega^c = \Omega^0 \cup \Omega^1$. We will also make frequent use of the following abbreviations. We use '$|\varphi|$' to refer to the proof set of φ in Ω^c; i.e., the set of all $\Gamma^n \in \Omega^c$ such that $\varphi \in \Gamma^n$. We use '$\mathfrak{B}(\mathbf{X})$' to refer to the closure of $\mathbf{X}$ under complements and binary intersections. And finally, for all i and Γ^n,

$$\Gamma^n \backslash A_i = \{\varphi : A_i\varphi \in \Gamma^n\}$$

$$P^{\star}_{\alpha,\Gamma^n} = \{\Delta^1 \in \Omega^c : \Gamma^n \backslash I_i \subseteq \Delta^1\}$$

'$\Gamma^n \backslash X_i$' and '$\Gamma^n \backslash I_i$' are defined in a similar fashion.

We can now characterise the canonical PWA model, M^c:

Definition 3 $M^c = (\Omega^c, \{\mathcal{N}^c_i\}_{i \in \mathbf{Ag}}, \{\mathcal{A}^c_i\}_{i \in \mathbf{Ag}}, \pi^c)$, where:

(1) $\Omega^c = \Omega^0 \cup \Omega^1$

(2) $\mathcal{N}^c_i(\Gamma^n) = \begin{cases} \varnothing, & \text{if } \Gamma^n \backslash A_i = \varnothing \\ \{|\varphi| : X_i\varphi \in \Gamma^n\} \cup \{\Omega^1, \Omega^0\}, & \text{if } \Gamma^n \backslash A_i \neq P^{\star}_{\alpha,\Gamma^n} = \varnothing \\ \{|\varphi| : X_i\varphi \in \Gamma^n\} \cup \{P^{\star}_{\alpha,\Gamma^n}\}, & \text{if } P^{\star}_{\alpha,\Gamma^n} \neq \varnothing \end{cases}$

[17] Thanks are due to Jessica Isserow, Harvey Lederman, and two anonymous referees for discussion and comments. This project has received funding from the European Unions Horizon 2020 research and innovation programme under the Marie Skodowska-Curie grant agreement No 703959.

(3) $\mathcal{A}^c_i(\Gamma^n) = \begin{cases} \varnothing, & \text{if } \Gamma^n \backslash A_i = \varnothing \\ \mathfrak{B}(\{|\varphi| : A_i\varphi \in \Gamma^n\} \cup 2^{\Omega^1}), & \text{if } \Gamma^n \backslash A_i \neq \varnothing \end{cases}$

(4) $\pi^c(p) = |p|$

We will also need a few basic lemmas.

Lemma 1 *For all* $\varphi, \psi \in \mathcal{L}^{AXI}(\Phi)$,

1. $|\neg\varphi| = \Omega^c \setminus |\varphi|$
2. $|\varphi \wedge \psi| = |\varphi| \cap |\psi|$
3. $|\top| = \Omega^c$

Proof. The proof is no different than for standard canonical models where $\Omega^c = \{\Gamma : \textsc{Max}\Gamma\}$, and therefore omitted. □

Lemma 2 $\mathfrak{B}(\{|\varphi| : A_i\varphi \in \Gamma^n\}) = \{|\varphi| : A_i\varphi \in \Gamma^n\}$

Proof. All we are trying to show here is that $\{|\varphi| : A_i\varphi \in \Gamma^n\}$ by itself is closed under complementation and binary intersections. This will be useful for the next lemma.

That $\{|\varphi| : A_i\varphi \in \Gamma^n\}$ is closed under complementation: Suppose that $P \in \{|\varphi| : A_i\varphi \in \Gamma^n\}$. So, $P = |\varphi|$, for some formula φ such that $A_i\varphi \in \Gamma^n$. By SYM, it follows that $A_i\neg\varphi \in \Gamma^n$. Hence, $|\neg\varphi| \in \{|\varphi| : A_i\varphi \in \Gamma^n\}$, and by Lemma 1 (1), $|\neg\varphi| = \Omega^c \setminus |\varphi|$. That $\{|\varphi| : A_i\varphi \in \Gamma^n\}$ is closed under binary intersections follows a basically similar structure, using CON and Lemma 1 (2). □

Lemma 3 $|\varphi| \in \mathfrak{B}(\{|\varphi| : A_i\varphi \in \Gamma^n\} \cup 2^{\Omega^1})$ *just in case* $|\varphi| \in \{|\varphi| : A_i\varphi \in \Gamma^n\}$, *whenever* $\Gamma^n \backslash A_i \neq \varnothing$

Proof. The right-to-left direction is trivial. To establish the left-to-right direction, we suppose throughout that $\Gamma^n \backslash A_i \neq \varnothing$, and hence $\{|\varphi| : A\varphi \in \Gamma^n\} \neq \varnothing$. We first show that if a proposition P belongs to $\mathfrak{B}(\{|\varphi| : A_i\varphi \in \Gamma^n\} \cup 2^{\Omega^1})$, then P satisfies at least one of the following three conditions:

c_1 $\quad P \in \{|\varphi| : A_i\varphi \in \Gamma^n\}$

c_2 $\quad P \subseteq \Omega^1$

c_3 $\quad P \cap \Omega^0 = P' \cap \Omega^0$, for some $P' \in \{|\varphi| : A_i\varphi \in \Gamma^n\}$

It's trivial that if $P \in \{|\varphi| : A_i\varphi \in \Gamma^n\} \cup 2^{\Omega^1}$, then P satisfies c_1 or c_2. So we only need to show that closing $\{|\varphi| : A_i\varphi \in \Gamma^n\} \cup 2^{\Omega^1}$ under complements and binary intersections will not leave us with any P which don't satisfy (at least) one of c_1, c_2, or c_3. We'll proceed in two steps.

For the first step, note that:

(a) If P satisfies c_1, then $\Omega^c \setminus P$ satisfies c_1

(b) If P satisfies c_2 or c_3, then $\Omega^c \setminus P$ satisfies c_3

(a) follows immediately from Lemma 2. For (b), note first that if P satisfies c_2, then $P \cap \Omega^0 = \varnothing$, which belongs to $\{|\varphi| : A_i\varphi \in \Gamma^n\}$ and therefore satisfies c_3 trivially. Furthermore, supposing that P satisfies c_3, P includes exactly those states of Ω^0 which belong to some $|\varphi|$. So, $(\Omega^c \setminus P) \cap \Omega^0$ includes just those states of Ω^0 which aren't included in P, which are exactly those in $|\neg\varphi| \cap \Omega^0$. Next,

(c) If P_1 satisfies c_2, then $P_1 \cap P_2$ satisfies c_2 (for any P_2)
(d) If P_1 and P_2 both satisfy c_1, then $P_1 \cap P_2$ satisfies c_1
(e) If P_1 and c_2 both satisfy c_3, then $P_1 \cap P_2$ satisfies c_3
(f) If P_1 satisfies c_1, and P_2 satisfies c_3, then $P_1 \cap P_2$ satisfies c_3

(c) is obvious, and (d) is an immediate consequence of Lemma 2. Similarly, (e) and (f) follow more or less directly from Lemma 2, which entails that if $P, P' \in \{|\varphi| : A_i\varphi \in \Gamma^n\}$, then $P \cap P' \cap \Omega^0 = P'' \cap \Omega^0$, for some $P'' \in \{|\varphi| : A_i\varphi \in \Gamma^n\}$.

Given facts (a) through (f) plus the fact that every $P \in \{|\varphi| : A_i\varphi \in \Gamma^n\} \cup 2^{\Omega^1}$ satisfies either c_1 or c_2, we know that for any $P \in \mathfrak{B}(\{|\varphi| : A_i\varphi \in \Gamma^n\} \cup 2^{\Omega^1})$, P satisfies c_1, c_2, or c_3.

So now let $\boldsymbol{\mathcal{P}}$ refer to the set of all proof sets. We now prove that if $P \in \mathfrak{B}(\{|\varphi| : A_i\varphi \in \Gamma^n\} \cup 2^{\Omega^1})$ then $P \in \boldsymbol{\mathcal{P}}$ only if $P \in \{|\varphi| : A_i\varphi \in \Gamma^n\}$, which gets us to the left-to-right direction of the present lemma. By definition, if P satisfies c_1, then it belongs to $\{|\varphi| : A_i\varphi \in \Gamma^n\}$. Likewise, if P satisfies c_2, then $P \in \boldsymbol{\mathcal{P}}$ only if $P = \varnothing$, in which case it also belongs to $\{|\varphi| : A_i\varphi \in \Gamma^n\}$. And finally, if P satisfies c_3, then $P \in \boldsymbol{\mathcal{P}}$ only if $P \in \{|\varphi| : A_i\varphi \in \Gamma^n\}$. For c_3 implies that P, whatever it is, contains exactly those states of Ω^0 which are also contained in some $P' \in \{|\varphi| : A_i\varphi \in \Gamma^n\}$. Suppose then that $P \in \boldsymbol{\mathcal{P}}$. Then, $\Gamma^1 \in P$ iff $\Gamma^0 \in P$, from which it follows that $P = P'$ for some $P' \in \{|\varphi| : A_i\varphi \in \Gamma^n\}$. □

With that out of the way, we first want to show that the canonical model so-characterised actually belongs to the class of PWA models, $\boldsymbol{\mathcal{M}}$:

Lemma 4 $M^c \in \boldsymbol{\mathcal{M}}$

Proof. That $\mathcal{A}_i^c$ and $\mathcal{N}_i^c$ satisfy *(pla)*: For the case when $\mathcal{A}_i^c = \varnothing$, $\mathcal{N}_i^c = \varnothing$ by definition. For the cases where $\mathcal{A}_i^c$ is non-empty, note that when $P^\star_{\alpha,\Gamma^n} = \varnothing$, then Ω^1 and Ω^0 are in both $\mathcal{N}_i^c(\Gamma^n)$ and $\mathcal{A}_i^c(\Gamma^n)$; and when $P^\star_{\alpha,\Gamma^n} \neq \varnothing$, $P^\star_{\alpha,\Gamma^n} \in \mathcal{N}_i^c(\Gamma^n)$ and (trivially) $P^\star_{\alpha,\Gamma^n} \in \mathcal{A}_i^c(\Gamma^n)$. So, it suffices to show that

$$\{|\varphi| : X_i\varphi \in \Gamma^n\} \subseteq \{|\varphi| : A_i\varphi \in \Gamma^n\}$$

This follows from XI and IA. For suppose that $B\varphi \in \Gamma^n$. Then $\vdash_\Sigma X_i\varphi \rightarrow (I_i\varphi \rightarrow A_i\varphi)$. So $A_i\varphi$ is derivable from $X_i\varphi$ in Σ, hence $A_i\varphi \in \Gamma^n$. Hence, $|\varphi| \in \{|\varphi| : X_i\varphi \in \Gamma^n\}$ only if $|\varphi| \in \{|\varphi| : A_i\varphi \in \Gamma^n\}$.

That $\mathcal{A}_i^c$ satisfies *(sym)* and *(con)* is true by construction; that $\mathcal{A}_i^c$ and $\mathcal{N}_i^c$ satisfy *(nax)* is straightforward and left to the reader. □

The next step is to establish a truth lemma:

Lemma 5 $M^c, \Gamma^n \models \varphi$ *iff* $\varphi \in \Gamma^n$

Proof. The proof proceeds by induction on the complexity of the formula φ. For the cases where $\varphi = p$, $\varphi = \neg\psi$, and $\varphi = \psi \wedge \gamma$, the argument is standard. Hence we will only focus on the following three cases:

(a) $\varphi = X_i\psi$
(b) $\varphi = A_i\psi$
(c) $\varphi = I_i\psi$

For case (a): Under the inductive hypothesis, $\|\psi\|^{M^c} = |\psi|$, so $\|\psi\|^{M^c} \in \mathcal{N}_i^c(\Gamma^n)$ iff $|\psi| \in \mathcal{N}_i^c(\Gamma^n)$. Suppose that $M^c, \Gamma^n \models X_i\psi$. By Definition 2, it follows that $\|\psi\|^{M^c} = |\psi| \in \mathcal{N}_i^c(\Gamma^n)$. As before, let $\boldsymbol{\mathcal{P}}$ refer to the set of all proof sets. $\Omega^0 \notin \boldsymbol{\mathcal{P}}$ and $P^\star_{\alpha,\Gamma^n} \in \boldsymbol{\mathcal{P}}$ only if $P^\star_{\alpha,\Gamma^n} = \varnothing$ (since $\varnothing \subseteq P^\star_{\alpha,\Gamma^n} \subseteq \Omega^1$). So by Definition 3, if $|\psi| \in \mathcal{N}_i^c(\Gamma^n)$, then $|\psi| \in \{|\varphi| : X_i\varphi \in \Gamma^n\}$, and so $X_i\psi \in \Gamma^n$. In the other direction, if $X_i\psi \in \Gamma^n$, then $|\psi| = \|\psi\|^{M^c} \in \mathcal{N}_i^c(\Gamma^n)$, so $M^c, \Gamma^n \models X_i\psi$.

For case (b): Given Lemma 3, the argument that $M^c, \Gamma^n \models A_i\psi$ iff $A_i\psi \in \Gamma^n$ is exactly parallel to case (a).

For case (c): Note first that $P^\star_{\alpha,\Gamma^n} \subseteq |\varphi|$ for any $|\varphi| \in \{|\varphi| : X_i\varphi \in \Gamma^n\}$. (Recall that $P^\star_{\alpha,\Gamma^n} = \{\Delta^1 : \Gamma^n \backslash I_i \subseteq \Delta^1\}$, and by XI, $\Gamma^n \setminus X_i \subseteq \Gamma^n \backslash I_i$.) Given this and Definition 3,

$$\bigcap \mathcal{N}_i^c(\Gamma^n) = P^\star_{\alpha,\Gamma^n}$$

Given this and Definition 2, $M^c, \Gamma^n \models I_i\psi$ iff $\Gamma^n \in \|A_i\psi\|^{M^c}$ and $P^\star_{\alpha,\Gamma^n} \subseteq \|\psi\|^{M^c}$. So suppose that $I_i\psi \in \Gamma^n$. Given IA, we know then that $A_i\psi \in \Gamma^n$; so $\Gamma^n \in |A_i\psi|$, and by the inductive hypothesis, $\Gamma^n \in \|A_i\psi\|^{M^c}$. Furthermore, we know that $\psi \in \Delta^1$ for every $\Delta^1 \in P^\star_{\alpha,\Gamma^n}$, so $P^\star_{\alpha,\Gamma^n} \subseteq |\psi| = \|\psi\|^{M^c}$, and $P^\star_{\alpha,\Gamma^n} \subseteq \|\psi\|^{M^c}$. Hence, if $I_i\psi \in \Gamma^n$, then $M^c, \Gamma^n \models I_i\psi$.

For the other direction, suppose that $M^c, \Gamma^n \models I_i\psi$. By the points just established, it follows that $P^\star_{\alpha,\Gamma^n} \subseteq \|\psi\|^{M^c}$, and given the inductive hypothesis, $P^\star_{\alpha,\Gamma^n} \subseteq |\psi|$. From this it follows that the set $\Gamma^n \backslash I_i \cup \neg\psi$ is inconsistent. For, suppose that $\Gamma^n \backslash I_i \cup \neg\psi \subseteq \Lambda$, for some maximal consistent set Λ. It would follow that $M^c, \Lambda^1 \models \neg\psi$, so by Lemma 1, $\Lambda^1 \notin |\psi|$. However this cannot be, since $\Lambda^1 \in P^+_{\alpha,\Gamma^n} \subseteq |\psi|$.

Since $\Gamma^n \backslash I_i \cup \neg\psi$ is inconsistent, some finite set $\{\varphi_1, \ldots, \varphi_n, \neg\psi\} \subseteq \Gamma^n \backslash I_i \cup \neg\psi$ is inconsistent. Thus:

$$\vdash_\Sigma \varphi_1 \to (\ldots(\varphi_n \to \psi)\ldots)$$

Given N_{AX} and XI, we have $I_i(\varphi_1 \to (\ldots(\varphi_n \to \psi)\ldots)) \in \Gamma^n$ whenever $A_i(\varphi_1 \to (\ldots(\varphi_n \to \psi)\ldots)) \in \Gamma^n$.

Next, note that since $\varphi_1, \ldots, \varphi_n \in \Gamma^n \backslash I_i$, we also get that $I_i\varphi_1, \ldots, I_i\varphi_n$ are in Γ^n. Under IA, it follows that $A_i\varphi_1, \ldots, A_i\varphi_n$ are also in Γ^n. Furthermore, note that by Definition 2, $\Gamma^n \in \|A_i\psi\|^{M^c}$, which given the earlier points means that $A_i\psi \in \Gamma^n$. Finally, for any pair of formulas φ and γ, if $A_i\varphi \in \Gamma^n$ and $A_i\gamma \in \Gamma^n$, then by applications of SYM and CON, $A_i\neg(\varphi \wedge \neg\gamma) \in \Gamma^n$. Since we can also show that awareness is closed under logical equivalence, so $A_i(\varphi \to \gamma) \in \Gamma^n$.

Putting the points of the previous paragraph together, in Γ^n we have:

$A_i\psi$,
$A_i(\varphi_n \to \psi)$,
$A_i(\varphi_{n-1} \to (\varphi_n \to \psi))$
$\vdots$
$A_i(\varphi_1 \to (\ldots(\varphi_n \to \psi)\ldots))$

Now, by finitely many applications of K_{AI}, we can derive that $I_i\psi \in \Gamma^n$. Hence, if $M^c, \Gamma^n \models I_i\psi$, then $I_i\psi \in \Gamma^n$. □

We can then apply a standard argument to get from Definition 3, Lemma 4 and Lemma 5 to establish Theorem 1. See (Chellas 1980, pp. 59ff) for details.

Appendix B

For Theorem 3, I'll just sketch a proof that $\Sigma \cup \{\text{ARX}\}$ is complete for the class of PWA models which satisfy (*arx*); the full proof proceeds along the same lines for the (*ara*) and (*ari*). We keep the characterisation of the canonical models M^c as given in Definition 3, with the obvious modification that Ω^c is now composed of maximal consistent sets relative to $\Sigma \cup \{\text{ARX}\}$. The proof is then the same as for Theorem 1, with the following addition to Lemma 4:

That M^c satisfies (*arx*): Let $Q = \{\Delta^i : P \in \mathcal{N}^c_j(\Delta^i)\}$. We need to show that if $Q \in \mathcal{A}^c_i(\Gamma^n)$, then $P \in \mathcal{A}^c_i(\Gamma^n)$. Assume that $Q \in \mathcal{A}^c_i(\Gamma^n)$. Now, either $P \notin \boldsymbol{\mathcal{P}}$, or $P \in \boldsymbol{\mathcal{P}}$. If the former, then by Definition 3, either $P \subseteq \Omega^1$ or $P = \Omega^0$; in either case, $P \in \mathcal{A}^c_i(\Gamma^n)$, so (*arx*) is straightforwardly satisfied whenever $P \notin \boldsymbol{\mathcal{P}}$.

Suppose then that $P \in \boldsymbol{\mathcal{P}}$. Since Definition 3 implies in general that $\mathcal{N}^c_j(\Delta^1) = \mathcal{N}^c_j(\Delta^0)$, so $\Delta^1 \in Q$ iff $\Delta^0 \in Q$. By points already established (Lemma 3), then, $Q \in \mathcal{A}^c_i(\Gamma^n)$ only if $Q \in \{|\varphi| : A_i\varphi \in \Gamma^n\}$; so $Q = |\varphi|$ for some φ such that $A_i\varphi \in \Gamma^n$. Furthermore, since by hypothesis $P \notin (2^{\Omega^1} \setminus \{\varnothing\}) \cup \{\Omega^0\}$, P will be in $\mathcal{N}^c_j(\Delta^i)$ if and only if $P = |\psi|$ for some ψ such that $X_j\psi \in \Delta^i$. So $Q = |X_j\psi|$, for some ψ. Putting these two facts together, we know that $A_iX_j\psi \in \Gamma^n$; and by ARX, $A_i\psi \in \Gamma^n$; so $|\psi| \in \{|\varphi| : A_i\varphi \in \Gamma^n\}$, which is of course a subset of $\mathcal{A}^c_i(\Gamma^n)$.

References

Chalmers, D. (2002). On sense and intension. *Philosophical Perspectives 16*, 135–182.

Chalmers, D. (2006). *The Foundations of Two-Dimensional Semantics*, pp. 55–140. Oxford University Press.

Chalmers, D. (2011). *The Nature of Epistemic Space*, pp. 60–107. Oxford: Oxford University Press.

Chellas, B. F. (1980). *Modal Logic: An Introduction.* Cambridge: Cambridge University Press.

Chen, Y.-C., J. C. Ely, and X. Luo (2012). Note on unawareness: Negative introspection versus AU introspection (and KU introspection). *International Journal of Game Theory 41*, 325–329.

Cozic, M. (2016). Probabilistic unawareness. *Games 7*(4), 38.

Dekel, E., B. L. Lipman, and A. Rustichini (1998). Standard state-space models preclude unawareness. *Econometrica 66*, 159–173.

Egan, A. (2008). Seeing and believing: perception, belief formation and the divided mind. *Philosophical Studies 140*(1), 47–63.

Elga, A. and A. Rayo (2015). Fragmentation and information access.

Elliott, E., K. McQueen, and C. Weber (2013). Epistemic two-dimensionalism and arguments from epistemic misclassification. *Australasian Journal of Philosophy 91*(2), 375–389.

Fagin, R. and J. Y. Halpern (1988). Belief, awareness, and limited reasoning. *Artificial Intelligence 34*, 39–76.

Fritz, P. and H. Lederman (2015). Standard state space models of unawareness. In *Proceedings of TARK 2015.*

Halpern, J. Y. (2001). Alternative semantics for unawareness. *Games and Economic Behavior 37*, 321–339.

Halpern, J. Y. and L. C. Rêgo (2009). Reasoning about knowledge of unawareness. *Games and Economic Behavior 67*(2), 503–525.

Heifetz, A., M. Meier, and B. C. Schipper (2006). Interactive unawareness. *Journal of Economic Theory 130*, 78–94.

Heifetz, A., M. Meier, and B. C. Schipper (2008). A canonical model for interactive unawareness. *Games and Economic Behavior 62*, 304–324.

Hintikka, J. (1962). *Knowledge and Belief: An introduction to the logic of the two notions*. Ithaca: Cornell University Press.

Hofweber, T. (2016). *Are there ineffable aspects of reality?* Oxford: Oxford Unviersity Press.

Jackson, F. (1998). *From metaphysics to ethics*.

Jackson, F. (2009). *Possibilities for representation and credence: two space-ism versus one space-ism*. Oxford: Oxford University Press.

Lewis, D. (1982). Logic for equivocators. *Nous 16*(3), 431–441.

Lewis, D. (1986). *On the Plurality of Worlds*. Cambridge University Press.

Li, J. (2009). Information structures with unawareness. *Journal of Economic Theory 144*, 977–993.

Modica, S. and A. Rustichini (1994). Awareness and partitional information structures. *Theory and Decision 37*(1), 107–124.

Modica, S. and A. Rustichini (1999). Unawareness and partitional information structures. *Games and Economic Behavior 27*, 265–298.

Montague, R. (1970). Universal grammar. *Theoria 36*, 373–398.

Nagel, T. (1986). *The view from nowhere*. Oxford University Press.

Schipper, B. C. (2013). Awareness-dependent subjective expected utility. *International Journal of Game Theory 42*, 725–753.

Schipper, B. C. (2014). Preference-based unawareness. *Mathematical Social Sciences 70*, 34–41.

Schipper, B. C. (2015). *Awareness*, pp. 77–146. London: College Publications.

Scott, D. (1970). *Advice in Modal Logic*. Reidel.

Sillari, G. (2008). Quantified logic of awareness and impossible possible worlds. *The Review of Symbolic Logic 1*(4), 514–529.

Stalnaker, R. (1978). *Assertion*, Volume 9, pp. 7895. New York: New York Academic Press.

Stalnaker, R. C. (1984). *Inquiry*. London: The MIT Press.

Stalnaker, R. C. (1991). The problem of logical omniscience, I. *Synthese 89*(3), 425–440.

Velázquez-Quesada, F. R. (2013). Explicit and implicit knowledge in neighbourhood models. In G. D., R. O., and H. H. (Eds.), *International Workshop on Logic, Rationality and Interaction*, pp. 239–252. Springer.

Walker, O. (2014). Unawareness with "possible" possible worlds. *Mathematical Social Sciences 70*, 23–33.