Journal of Philosophical Logic manuscript No.
(will be inserted by the editor)

Unawareness and Information Access in Possible Worlds Models of Belief

First Author

Received: date / Accepted: date

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 in principle access given their explicit beliefs. In this paper, I develop a model of explicit belief, awareness, and informational content, along with an sound and complete axiomatisation. I furthermore defend the model against the seminal impossibility result of Dekel, Lipman and Rustichini, according to which three intuitive conditions preclude non-trivial unawareness on any possible worlds model of belief.

Keywords Unawareness · Epistemic logic · Information access · Possible worlds semantics · Neighbourhood structures

1 Introduction

Iggy is an agent, who may or may not be ideally rational. In his capacity as an agent, Iggy must have at least some of what we'll refer to broadly as *epistemic attitudes* regarding some range of propositions. There will be some propositions that Iggy knows, some he believes, some he disbelieves, and some that he's contemplated but he's not sure about. Question: for *any* proposition φ, must Iggy have some epistemic attitude or other regarding φ?

Most will want to say that he need not. To know, believe, be uncertain about, or even suspend judgement regarding a proposition requires at least the capacity to *entertain* that proposition, to *represent* it some way or another in thought—and there may be some propositions that Iggy has never entertained, will never entertain, and perhaps *cannot* even entertain.

This project has received funding from [*anonymous*].

F. Author
first address
Tel.: +123-45-678910
Fax: +123-45-678910
E-mail: fauthor@example.com

For example, suppose that Iggy belongs to an isolated tribe located deep in a far off jungle, which has had no contact with any other cultural group for hundreds of years. Most people wouldn't want to say that Iggy knows, believes, or is uncertain about the proposition that we would ordinarily 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 Iggy is agnostic, or has suspended judgement, on the matter of eastern grey kangaroo habitats. Suspension of judgement is typically understood as a kind of mental act that one might perform upon entertaining some proposition, when one decides that there's insufficient reason to believe one way or the other. But Iggy might not even know that there's a proposition here for him to suspend judgement about. Indeed, he may not have the basic mental resources necessary to even think the thought, let alone decide whether he has sufficient evidence to believe it.

So let us admit, at least in principle, that there may be some propositions of which Iggy is unaware. I use this phrase semi-stipulatively: to be *unaware* of a proposition φ is to lack the basic representational resources required to entertain φ-thoughts or otherwise have what were above called epistemic attitudes with φ as their content. To be *aware* of φ is to not be unaware of it.[1]

If this kind of *unawareness* exists, and is widespread, then it seems to be the kind of thing to which any developed formal theory of the epistemic attitudes ought to pay attention. Consequently, after saying more in the next section about what I mean by 'awareness' and why we should care about it, in Section 3 I will characterise a class of models—Possible Worlds Awareness (PWA) models—aimed at representing differences in states of awareness and unawareness, and how these interact with the epistemic attitudes and questions of information access. I focus on modelling the interaction of unawareness with the attitudes of belief and knowledge, though much of my discussion will also be of direct relevance to Bayesian models of uncertainty under unawareness (see, e.g., [2], [10]).

As the name suggests, PWA models adopt the kind of coarse-grained approach to propositions that's generally associated with possible worlds semantics. The models are therefore subject to the influential triviality result of Dekel, Lipman and Rustichini [14]. In Section 4, I describe Dekel *et al.*'s result and some natural extensions thereof, while in Sections 5–8, I critically discuss the intuitions and implicit philosophical assumptions that motivate it. As we will see, the force of their result depends much on the intended interpretation of the model, and (moreover) on questions regarding the nature and granularity of mental content.

2 Background

Before anything else, let me first say more about how 'awareness' is understood in the wider literature and how I will be using the term (Section 2.1), and why we should care about awareness so-understood (Section 2.2).

[1] It is worth emphasising the distinction in English between being aware *that* something is the case, and being aware *of* some thing. The former is synonymous with knowing a true proposition: Iggy is *aware that* φ iff φ is true and Iggy knows it; Iggy is *unaware that* φ iff φ is true and Iggy doesn't know it. On the other hand, Iggy would be aware *of* a proposition in something like the way he might be aware of an idea, an argument, or an oncoming train: by consciously attending to (a representation of) the object of one's awareness, or at least being in a position to do so.

2.1 What is awareness?

In their seminal work on the topic, Fagin and Halpern [16] were intentionally ambiguous with their use of 'awareness'. For some of their discussion, the term is used in connection with a notion of entertainability that's at least similar to that which I characterised above. However, they also use it to mark a distinction between the beliefs that a subject is consciously attending to versus her background beliefs; and between those possibilities that the agent has considered and incorporated into her deliberations versus those she's not yet considered or perhaps considered but forgotten about. And in still other cases, Fagin and Halpern use 'awareness' and 'unawareness' to distinguish between those conclusions which can and cannot be derived from a 'database' of stored beliefs within a given amount of time.

In other words, we have had from the start four distinct (though by no means unrelated) notions all being discussed under the same heading. Roughly, we have awareness of φ as:

Entertainability: Being able to entertain the possibility that φ.
Attentional: Consciously attending to the possibility that φ.
Deliberative: Having considered the possibility that φ in one's reasoning.
Derivability: Being able to derive that φ within a given time.

All four admit of differences in degree and further precisifications. For example, some possibilities might be more or less at the forefront of Iggy's conscious attention, and some things might be more or less easy to derive, or derivable only given certain conditions.

Unfortunately, matters have not in general become clearer since Fagin and Halpern's essay. Awareness is often characterised in the first instance as a "lack of ability to conceive", or a "lack of concept" (e.g., in [24], [43], and [55]), which perhaps suggests that something like the entertainability sense is intended. Moreover, many of the formal conditions that awareness is standardly taken to satisfy seem to fit best with something like the entertainability sense, whereas they fit relatively poorly with the attentional, deliberative, and derivability senses unless strong assumptions are made about the capacities of the agent in question. (See, e.g., SYM, CON_A, and especially the *AGPP* principle; these will be discussed in Section 3.) However, characterisations of 'awareness' in the literature tend to be perfunctory at best, and detailed exposition on the matter is rare and sometimes conflicting, making confident interpretation difficult.[2]

I note all this in order to draw attention to the fact that I will be understanding 'awareness' in this paper in a rather specific way—*viz.*, as a particular instance of the entertainability notion. This will be important throughout my discussion, and I certainly do not want to claim that all of my conclusions apply with equal force to every understanding of 'awareness' present in the literature. Hence, it will be useful to get clearer on just what 'awareness' means in the present context.

The following is a good rough-and-ready way to picture the kind of thing that I have in mind. Suppose we adopt a very naïve version of the Language of Thought theory combined with the Representational Theory of Mind, at least as applied to the epistemic attitudes (see [40] for details). We assume that for Iggy, and any

[2] For exceptions to this rule, see [20] and [48]. An orthogonal distinction in kinds of awareness, between awareness that attaches to propositions versus awareness that attaches to sentences or theorems, is also discussed in [20].

agent like him, mental representation is underwritten by an essentially linguistic system of meaningful symbols. Moreover, this symbolic 'language' looks a lot like a simplified version of English. Somewhere inside his head Iggy has, or is capable of 'uttering', various word-like representations which might denote individuals, properties, and relations, as well as perhaps operators, quantifiers, connectives, and so on. Call these his *concepts*. Iggy's concepts can be composed in a systematic way into sentence-like structures, which have the propositional contents that they do as a consequence of the meanings of their constituent concepts and the ways that those concepts are arranged. For Iggy to have any epistemic attitudes whatsoever regarding a proposition φ, he must be psychologically related to some sentence composed out of concepts he possesses which means that φ (cf. [17], [18]). Presumably, different agents will tend to have different concepts—different mental vocabularies, as it were. Iggy lacks any concepts that pick out kangaroos, for example, whereas the average Australian would typically possess such a concept. Call the concepts that Iggy does have his *basic representational resources*.

If we were to accept all this, then we could say that Iggy is *aware of* φ, in the sense that I have in mind, just in case he possesses concepts sufficient to compose a sentence that means that φ. Thus, he need not be presently and consciously attending to φ in order to be aware of it, and the entertaining itself might require significant mental effort—in fact it may even be physically impossible. So, for example, Iggy may be aware of the proposition expressed by the near-infinitely long English sentence 'Iggy believes that Jiggy believes that Iggy believes that... that Jiggy believes that φ', merely by being able to represent *himself*, his colleague *Jiggy*, φ and the *belief* relation—even if actually 'uttering' the relevant sentence in his Language of Thought would require many more neurons than could fit in any ordinary-sized human brain.

Now, I don't myself attach much credence to that picture of how the mind works. Even if mental representation *is* typically language-like in the way that the picture supposes, it's highly doubtful that the content of every epistemic attitude that Iggy has will itself be the meaning of some sentence-like symbol stored somewhere in his head (see esp. [11], [12]). (Not that advocates of the Language of Thought and/or the Representational Theory of Mind have ever thought otherwise—I called it "naïve" for a reason.) Furthermore, the picture gives the potentially misleading impression that the possession of certain very basic logical concepts is a mere contingent matter, a question of whether the agent has a particular word in their mental lexicon. Had Iggy merely lacked the concepts for *conjunction* and *negation*, say, then even if he could entertain φ and ψ individually, he might still have been unable to entertain $\neg\varphi$ and $\varphi \wedge \psi$.

Compare, for example, an account where the format of our mental representations is more map-like in nature (cf. [4], [34]). A map that represents that *Sydney is northeast of Melbourne* and that *Sydney is northeast of Canberra* automatically represents that *Sydney is northeast of Melbourne and Canberra*, and it doesn't need a special symbol denoting *conjunction* to do so—the propositional contents of maps are closed under conjunctions just as a consequence of the way that maps represent the world in general. Similarly, a map-like representational system that can represent that *Sydney is on the coast* is, by that very fact, able to represent that *Sydney is not on the coast* (e.g., simply move whatever represents the city inland); and if 'X' marks the spot where Blackbeard's treasure is, then the absence of an 'X' marks the spots where it isn't. Maps don't usually have symbols on them

that represent *negation* or *absence*, but they still manage to tell us a lot about what is and *isn't* the case without them. Hence, Iggy's ability to entertain $\neg\varphi$ and $\varphi \wedge \psi$ *might* be an immediate consequence of his ability to entertain φ and ψ, due to the format that his mental representations take and the operation of the systems that use them, rather than being dependent on the presence of some special mental symbols that specifically denote logical operators and connectives. (I'll say more in relation to this in Section 3.1)

So the naïve picture isn't perfect, but it does manage to capture the core idea of what I mean by 'awareness' well enough, and it does so in a relatively intuitive way. In general: Iggy can represent different ways the world might be. Presumably, his capacity to represent those different ways depends on his having some stock of basic representational resources that can be systematically restructured or recombined to represent new possibilities (as argued in [18], [19]). This seems highly plausible regardless of the exact format that our mental representations take, map-like or sentence-like or otherwise. Some of those possibilities that Iggy has the basic resources to represent he never will in fact represent, and some may be too complex or time-consuming to actually represent given his limited cognitive resources—but there will still be a good sense in which he possesses the *basic representational resources* required to do so. On the other hand, some propositions Iggy will be unable to represent by virtue of lacking the requisite representational resources; some distinctions between ways the world might be won't even be on Iggy's mental radar. Of these, he is unaware.

2.2 Why should we care about awareness?

I am inclined to think that all four senses of 'awareness' are important, each in their own way and to varying extents. So why should we care about the particular kind of awareness *qua* entertainability that I've specified?

Here's one reason: it is the weakest of the four senses. One gets to be aware of φ in the attentional, deliberative, or derivability senses only if one is aware of φ in the entertainability sense. Moreover, awareness in any of the other three senses will tend to vary dependent on the rational capacities of the agent in question. To put the point roughly: amongst those propositions that Iggy can in principle entertain, he will be able to consciously attend to more, incorporate more into his reasoning, and derive more faster, the bigger and better his brain is. What an agent can entertain in principle thus sets bounds on their awareness in the other senses, bounds which depend only on that agent's basic representational resources (and not on their capacity for reasoning, attentional resources, etc.).

Furthermore, if unawareness in this sense is widespread,[3] then it is an important phenomenon that ought to be represented in our best models of the epistemic attitudes—including (but not limited to) our models of ideally rational agents. Consider, for instance, the standard possible worlds models of belief in the style of Hintikka [25]. We begin with a rich space of possible worlds, Ω, subsets of which represent the contents of our attitudes.[4] Our agent Iggy (i) can then be associated with a binary relation R_i on Ω which for each world ω (effectively) picks out a

[3] See Section 6 for discussion on how widespread we should expect 'awareness' to be.

[4] By my use of 'possible worlds,' I mean to exclude specifically worlds which are either not maximally specific, or inconsistent with classical propositional logic. Much of what I say will

proposition $R_i(\omega)$ containing all and only those worlds that Iggy considers possible at ω. We then say that Iggy *believes* that φ at ω iff φ is true at every world in $R_i(\omega)$. Problem: if $R_i(\omega)$ is even the least bit specific, then a vast number of things will be true at every world in $R_i(\omega)$, and we can't expect Iggy to believe *all* of them—and not just because he may not be a very good deductive reasoner. Make Iggy as logically gifted as you like, give him unlimited working memory, and all the time in the world; still, if he can't even entertain certain propositions then it's wrong to say that he believes them.

A simple example of what I mean. Suppose that there are only two atomic propositions, φ and ψ, and exactly four possible worlds ω_1, ω_2, ω_3, ω_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 $R_i(\omega_1) = \{\omega_1, \omega_2\}$. This is just the set of worlds where φ is true, so we can say that Iggy believes φ. Now presumably, if Iggy believes φ then he can entertain φ; and if he can entertain φ then he can distinguish between φ and $\neg\varphi$. So it looks fair to say that Iggy is able to entertain $\neg\varphi$ at ω_1. Likewise, since he can entertain φ and $\neg\varphi$, he can entertain $\varphi \wedge \neg\varphi$ and $\varphi \vee \neg\varphi$. The latter is true at every world in $R_i(\omega_1)$, and it's reasonable enough to expect that Iggy believes $\varphi \vee \neg\varphi$. At the very least, he can certainly reason his way to that conclusion given his other beliefs without much difficulty.

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

For the sake of clarification, it will be helpful to discuss two very common responses to this kind of example before we move on.

1. First, you may have the thought that it would be equally wrong to say that Iggy is able to entertain $\psi \vee \neg\psi$, even though $\psi \vee \neg\psi$ picks out just the same set of possible worlds that $\varphi \vee \neg\varphi$ does; thus the possible worlds model of content is inadequate for understanding states of unawareness in any case.

If you do have that thought, then you will probably also think that possible worlds theories of content suffer from problems relating to hyperintensionality more generally. And you may well might be right. (I'll consider this matter again in Section 8.) But the phenomenon that I want to discuss in this paper is very general, and arises regardless of how granular you think mental content is. Even

not hang on any specific account of what 'worlds' are; though see Section 6 for discussion on some issues that depend on exactly which worlds constitute Ω.

if you accept a very coarse-grained conception of propositions—i.e., one such that $\varphi \vee \neg\varphi$ and $\psi \vee \neg\psi$ are by virtue of their classical logical equivalence the very same proposition just described or picked out in two different ways, or under different *modes of presentation*—then you will *still* need to deal with the possibility of unawareness. It's not plausible to say that Iggy believes $\varphi \vee \psi$ under *any* mode of presentation. (Henceforth, we'll refer to advocates of coarse-grained theories of content as *possible worlds theorists*.)

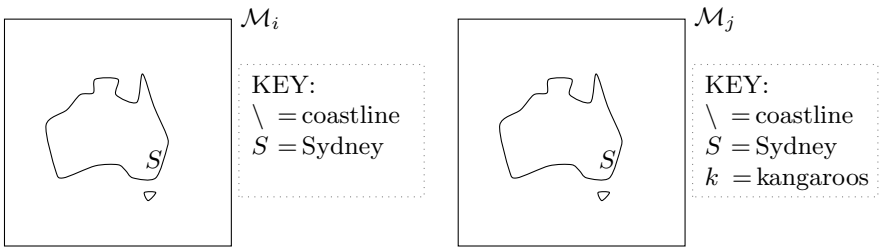
2. Second, you may think that the relational model has no real problems with unawareness at all—at least, not once it's interpreted *in the right way*. After all, it is frequently said that the relational model is best viewed as a way of representing an agent's *implicit* rather than *explicit* beliefs, where the former are understood to capture something like the informational content contained in the beliefs that the agent actually has 'stored' somewhere in her head. Thus, if we interpret the model this way, there need not be any issues whatsoever with claiming that Iggy *implicitly* believes both $\varphi \vee \psi$ and $\varphi \vee \neg\psi$, merely on the basis of the assumption that he *explicitly* believes φ. After all, the former are in a perfectly well-defined sense built into the informational content of the latter.

Now I want to be clear that I have no objections to interpreting the relational model like this; if that's the kind of thing you want to represent, then so be it. Nevertheless, in response to the point I do want to introduce a distinction that I think is very important when it comes to thinking about the *informational content* of our epistemic attitudes, and one which underscores the need for incorporating some representation of unawareness into our formal models.

Say that in a *broad* sense, the informational content of a set of (explicitly represented, or 'stored') epistemic attitudes consists exactly in anything and everything that's entailed by the contents of those attitudes, whether separately or in conjunction. It's in this sense that it would be unproblematic to say that Iggy's explicit 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 draw it out.

But Iggy is *not* one of those with the required representational resources. For him, some of that broad informational content is in a very strong sense *inaccessible*: it is not available for him to use in reasoning, inference and decision making (cf. [58]). For all rational purposes, that information is for him completely invisible. Iggy may not even know that there's a distinction that he's failed to consider, and no amount of reasoning with the distinctions that he does have will lead him to recognise what he's been missing. And information which is wholly invisible to Iggy is, from his own perspective, not really information at all. It seems clear then that there's an important theoretical role for a second, *narrow* notion of informational content—one that's specifically dependent upon awareness *qua* entertainability.

This point is, I think, especially straightforward if mental representation does turn out to be sentence-like: obviously, and regardless of how rational he is, Iggy cannot possibly derive from that which he explicitly believes any sentences the content of which can only be expressed using concepts that he lacks. But the distinction between broad and narrow informational content is still useful even if mental representations are map-like. Imagine, for example, that Iggy and Jiggy each has their own 'mental map', $\mathcal{M}_i$ and $\mathcal{M}_j$ respectively, which have slightly different keys but are in all other respects similar:



Though both maps explicitly represent that Australia is shaped like so, and that Sydney is located on its southeast coast, $\mathcal{M}_j$ represents a somewhat more specific possibility than $\mathcal{M}_i$ does: $\mathcal{M}_j$ explicitly represents that there are no kangaroos to be found anywhere within or around the country; whereas $\mathcal{M}_i$ is neutral on the matter of whether there are any kangaroos, and if so, where they might be located. But, more importantly, given the richer key associated with $\mathcal{M}_j$, Jiggy *could* represent the presence of kangaroos at any number of different locations, by simply adding a 'k' on the appropriate part of the map. Iggy's map, by contrast, lacks the resources to represent the presence or absence of kangaroos—draw a 'k' on any part of $\mathcal{M}_i$ and you're left with a meaningless squiggle, or perhaps a very odd coastline, and in Iggy's case the absence of a 'k' doesn't represent anything at all.

So now consider:

φ = Sydney is on the southeast coast of Australia.
ψ = There are kangaroos as far north as the Cape York Peninsula.

The 'explicit' contents of both Iggy's and Jiggy's mental maps entail that $\varphi \vee \psi$ and $\varphi \vee \neg\psi$, since they both clearly represent at least φ. We can all agree that $\varphi \vee \psi$ and $\varphi \vee \neg\psi$ belong to the broad informational contents of both maps. But Jiggy knows what it would take for $\varphi \vee \psi$ and $\varphi \vee \neg\psi$ to be true—he has the capacity to represent all the different ways those propositions might be true, he can imagine and entertain those possibilities and play around with them in his hypothetical reasoning. Iggy, by contrast, cannot: his map *entails* $\varphi \vee \psi$ and $\varphi \vee \neg\psi$, but he's not aware of these entailments. The fact that Iggy's mental map includes $\varphi \vee \psi$ and $\varphi \vee \neg\psi$ in its broad informational content is of no rational use to him even in principle, given his relatively limited conceptual resources.

So I take it that the distinction between broad and narrow informational content is theoretically well-motivated, and independent of questions concerning representational format. Furthermore, if we draw the natural corresponding distinction between *broad* and *narrow* implicit beliefs, then it's arguably the latter which fit better with the central function that the notion of 'implicit beliefs' were originally introduced to play—*viz.*, as a way of marking a distinction between that information which is explicitly represented by a thinker versus that which is not explicitly represented but is still *accessible for use* on the basis of that which is.[5]

But it doesn't really matter what words we use. What's important is whether the distinctions are useful—and the fact that an agent's *narrow implicit beliefs* have much closer ties to matters concerning the agent's rationality suggests that they are. If Iggy fails to act appropriately given his narrow implicit beliefs then he

[5] See [1], pp. 185ff, for an illuminating discussion of the notion of *implicit belief* in the context of sentence-like and map-like theories of mental representation.

is rationally criticisable, at least to some (possibly very small) extent. On the other hand, if Iggy fails to act appropriately given his non-narrow broad implicit beliefs, then any such criticism seems wrong-headed. One isn't the least bit epistemically blameworthy for lacking certain representational resources, and not having a way to mentally represent kangaroos doesn't make Iggy irrational under any remotely reasonable conception of rationality.

It's worth noting also that given a narrow understanding of informational content, there is still plenty of room for Stalnaker's [52] useful distinction between content which is *more or less* accessible for different kinds of applications. It's plausible to think that the way we represent information about the world makes a difference to how easily we might access it and reason with it on a given occasion, and consequently that at any time we probably only attend to a fragment of the total information we have represented across the full range of our epistemic attitudes. As Stalnaker notes, recognising the limits of attention and information access goes some of the way towards solving the problems of logical omniscience that all possible worlds theories of content have to deal with (cf. also [15], [31], [51], [58]). But any such a solution cannot be complete without also recognising the role that awareness *qua* entertainability has as a precondition for genuine informational access *simpliciter*.

In summary: awareness *qua* entertainability is theoretically important, as are the notions of narrow informational content and narrow implicit belief which directly depend upon it. Hence, in the next section I aim to outline a class of models that will distinguish appropriately between arbitrary agents' (i) states of awareness *qua* entertainability, (ii) their explicit knowledge and belief, and (iii) the narrow informational content that's in principle accessible given those explicit attitudes, all within the context of a possible worlds theory of content.

Ultimately, it would be nice to have a model that incorporates all of the different kinds of unawareness—variabilities in attentional awareness, for example, with perhaps a sliding scale for different degrees of informational accessibility and deliberative capacity, and multiple belief states at once to represent fragmentation. It would in any case be useful to have some representation of (i) what information an agent has access to in principle—what they could derive from their explicit beliefs if they were ideally rational—from (ii) that information that the agent does have access as a matter of fact, and (iii) that information that's readily available (or ready-to-degree-x available) to the agent given their bounded capacities. Each notion has a role to play in explaining action, reasoning, and ideal versus bounded rationality. But that is much more than I can hope for here, and an initial model for just one small part of that much larger picture will have to suffice here.

3 Possible Worlds Awareness

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. (Henceforth I will simply say 'awareness' rather than 'awareness *qua* entertainability'.) I will leave the interpretation of the X operators ambiguous for now, between an *explicit knowledge* and an *explicit belief* reading. The I operators are used to stand for either *narrow implicit knowledge* or *narrow implicit belief*, depending on how we choose to read the Xs. To make things easier, I'll focus just on the *belief* interpretation for this section and the next. I will have more to say on the different interpretations of X and I as we progress through later sections.

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 any given world ω. However, I will not assume that agents' beliefs satisfy any closure conditions except closure under necessary equivalence (which is essential to any possible worlds theory of content). Nor will I assume that any given agent is *ir*rational. I want the model to be flexible enough to allow for both rationality and irrationality of explicit belief. Thus, in characterising agents' explicit beliefs, I will use the strictly more general neighbourhood structures of [38] and [47].

A standard neighbourhood model M consists in (i) a set of worlds Ω, (ii) a set of what we will call *belief functions* $\mathcal{X}_i$ (one for each $i \in \mathbf{Ag}$) which map each world ω to a (potentially empty) set of (potentially empty) subsets of Ω, and (iii) a propositional valuation function $\pi : \Phi \mapsto 2^{\Omega}$. The satisfaction conditions for φ in the sub-language $\mathcal{L}^X(\Phi)$ 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 $\{\omega' : M, \omega' \models \varphi\} \in \mathcal{X}_i(\omega)$

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 $\|\varphi\|^M \in \mathcal{X}_i(\omega)$

Thus, we can use belief functions to characterise directly those propositions that we want to say i believes at a given world ω. Note that there is no inbuilt assumption, for any particular φ and any ω, that $\|\varphi\|^M$ must belong to $\mathcal{X}_i(\omega)$. Likewise, we don't assume that if $\|\varphi\|^M \in \mathcal{X}_i$ and ψ is true at every world in $\|\varphi\|^M$, then $\|\psi\|^M \in \mathcal{X}_i$. Indeed, the logic associated with neighbourhood models is very weak, consisting of just:

PROP. All classical propositional tautologies
MP. From φ and $\varphi \rightarrow \psi$, 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 [8], 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 states of awareness, which we will represent by an additional set

of *awareness functions*, $\mathcal{A}_i$. Like the belief 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 four additional conditions upon each agent's awareness function and its relationship with belief functions. I'll discuss these in turn.

3.1 Basic conditions on belief and awareness functions

pla. The most obvious condition concerns the relation between awareness and the explicit attitudes. An agent's awareness cannot float free of her explicit beliefs: if she isn't aware of a proposition, then she cannot explicitly believe it. I take it that this is analytic of 'awareness' if anything is, given how I've characterised the notion. For reasons that will become apparent below, in the axiom system we'll refer to this as PLA^a:

PLA^a. $\quad X_i\varphi \rightarrow A_i\varphi$

We capture PLA^a with the condition *pla*, that $\mathcal{X}_i(\omega) \subseteq \mathcal{A}_i(\omega)$, for all ω.

sym. Next, and a little less obviously, I will assume that awareness is closed under negations. So, if an agent is aware of φ, then she is aware of $\neg\varphi$. Hence,

SYM. $\quad A_i\varphi \rightarrow A_i\neg\varphi$

We ensure that this axiom holds by assuming that $\mathcal{A}_i$ is closed under complements (*sym*). In any classical logic SYM implies $A_i\varphi \leftrightarrow A_i\neg\varphi$.

I expect that SYM will be generally uncontroversial, and I have already mentioned some reasons to think that our *actual* states of awareness might be automatically closed under negations in Section 2.1. Nevertheless, allow me to here add some considerations in favour of SYM.

First: on any account of concepts, one possesses the concept for some kind F only if one is able to tell the Fs from the not-Fs. Concepts as they are widely understood either literally are, or otherwise form the representational bases for, our recognitional and sortal capacities; so, if you can think that *a is an F* you can think that *a is a not-F*, which to the possible worlds theorist just is to think that *a is not an F*. The same is plausibly true for propositions more generally. One can't really understand what it is for some possibility to be the case if one can't understand what it would take for it to not be the case. There is also a long and widespread tradition going back to [57]—one that's particularly common amongst possible worlds theorists—according to which one isn't able to represent the possibility that φ without being able to represent $\neg\varphi$. According to this tradition, to believe φ is just to carve the space of possibilities Ω down the line that divides φ and $\neg\varphi$, and toss the $\neg\varphi$ half away. So it's at least arguable, and quite widely assumed, that the capacity to believe anything seems to presuppose 'possession of the concept of negation', in at least the weak sense characterised by SYM.[6]

[6] Something similar holds on almost every account of uncertainty within the Bayesian tradition: having any uncertainty regarding φ implies having some uncertainty regarding $\neg\varphi$, while learning φ requires setting one's uncertainty regarding $\neg\varphi$ to zero.

con. We will also assume that awareness is closed under conjunctions: if an agent is aware of φ and ψ, then she is aware of $\varphi \wedge \psi$.

CON_A. $(A_i\varphi \wedge A_i\psi) \rightarrow A_i(\varphi \wedge \psi)$

Thus we will assume that the $\mathcal{A}_i$ are closed under binary intersections (*con*). In conjunction with SYM, this means that for all worlds ω and agents i, $\mathcal{A}_i(\omega)$ is either empty or a finitely additive Boolean algebra on Ω.

The reasons for accepting CON_A are not dissimilar from those in favour of SYM. Suppose that Iggy believes that φ, and believes that ψ. He may or may not therefore believe that $\varphi \wedge \psi$, but there can be hardly any doubt that he is *aware* of that possibility—in the sense that he has at least the basic representational resources required to represent that proposition. Even if those beliefs are stored in different 'fragments', there will be a good sense in which he *could* bring those beliefs together, and guide his behaviour on the basis of the information contained in the pair of them. As we've seen already, Iggy doesn't have to have a sentence in his head that specifically means $\varphi \wedge \psi$ in order to represent that proposition. It would be enough if he merely represents at once that the world is as φ says it is and that the world is as ψ says it is, and is able to draw conclusions as appropriate from that basis. This, I assume, is something that anyone with the capacity to represent φ and ψ individually has the requisite representational resources to do.

A clarificatory note: we will *not* be assuming that an awareness of $\varphi \wedge \psi$ implies an awareness of φ and awareness of ψ. In other words, we don't assume:

DIS_A. $A_i(\varphi \wedge \psi) \rightarrow (A_i\varphi \wedge A_i\psi)$

In any classical logic, having DIS_A alongside SYM and CON_A would quickly lead to triviality. As pointed out in [20], if an agent 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 then imply awareness of every ψ, which is clearly unacceptable. But the arguments just given in favour of CON_A don't also support DIS_A, and more generally it should be apparent that DIS_A is only plausible under the presupposition of a fine-grained theory of content. (Consider again the example from Section 2.2: Iggy is aware of φ, which just is $(\varphi \vee \psi) \wedge (\varphi \vee \neg\psi)$, yet Iggy isn't aware of either $\varphi \vee \psi$ or $\varphi \vee \neg\psi$.) Obviously, this is only a reason to reject DIS_A if we have good reasons to accept a coarse-grained conception of mental content—but that discussion will have to wait until Section 8.

nax. Finally, and mostly for the sake of simplicity, I will make one further assumption about the relationship between awareness and explicit belief:

N_{AX}. $A_i\top \rightarrow X_i\top$

In the context of the other axioms, this implies that 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 believes $\top$. Where 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 use of the condition *nax*, that if Ω belongs to $\mathcal{A}_i$, then it belongs to $\mathcal{X}_i$.

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 stronger) axiom N (discussed in Section 4). 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 $A_i\top \rightarrow I_i\top$. The soundness and completeness proofs will be left mostly unchanged.

3.2 PWA models

I will discuss further potential conditions on awareness and belief functions below, but for now let us move on to the basic PWA models:

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

1. Ω is a non-empty set
2. $\mathcal{X}_i$ and $\mathcal{A}_i$ are functions from Ω to 2^{2^Ω} satisfying (for all $\omega \in \Omega$):

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

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., $M, \omega \models \varphi$):

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 $\|\varphi\|^M \in \mathcal{X}_i(\omega)$
$M, \omega \models A_i\varphi$ iff $\|\varphi\|^M \in \mathcal{A}_i(\omega)$
$M, \omega \models I_i\varphi$ iff $\|\varphi\|^M \in \mathcal{A}_i(\omega)$ and $\bigcap \mathcal{X}_i(\omega) \subseteq \|\varphi\|^M$

Definition 2 should make it clear how the awareness functions work in relation to the belief functions. If i believes each of a collection of propositions $\varphi_1, \ldots, \varphi_n$ which jointly imply ψ, then we will say that i (narrowly) implicitly believes ψ whenever she's aware of ψ. This is modelled by first characterising (for each world ω) the largest set of worlds $P_1 \subseteq \Omega$ consistent with all of i's explicit attitudes at ω as the intersection of all the sets of worlds in $\mathcal{X}_i(\omega)$. We then say that i implicitly believes any P_2 such that $P_1 \subseteq P_2$ just in case i is aware of P_2. So, Iggy has the narrow implicit belief that φ just in case φ is something Iggy can entertain that's entailed by the conjunction of his explicit beliefs. Thus, $\mathcal{A}_i$ acts as a filter or sieve over the propositions that i implicitly believes in the broad sense, letting through only those that i implicitly believes in the narrow sense.

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$_A$. $(A_i\varphi \wedge A_i\psi) \to A_i(\varphi \wedge \psi)$
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 $\varphi \in \mathcal{L}^{AXI}(\Phi)$, $\boldsymbol{\mathcal{M}} \models \varphi$ if and only if $\vdash_\Sigma \varphi$.*

Proof See Appendix A.

Σ 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$_I$. $I_i\varphi \to \varphi$

In combination with XI, this will get us:

T$_X$. $X_i\varphi \to \varphi$

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

tx. If $P \in \mathcal{X}_i(\omega)$, then $\omega \in P$

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

Finally, it's worth noting that if we wanted to capture instead the other senses of 'awareness' outlined in Section 2.1, we might be able to do this by removing *pla*, *sym* and/or *con* from Definition 1 and modifying the axiom system appropriately. For example, I see no strong reasons to think that PLA^a and CON$_A$ are appropriate for either the attentional or deliberative senses of 'awareness', unless we make very strong idealising assumptions about the agent's capacities.

3.3 Relationship to other models

Before I discuss Dekel *et al.*'s triviality result and my response to it, in this subsection I'll briefly discuss the relationship between PWA models and other models in the literature. It should be emphasised that the notions of 'awareness' and 'implicit belief' referred to in the other works discussed herein need not be the specific kinds of awareness and implicit belief that I take to be represented by PWA models, though they are of course clearly related.

Considered primarily as a way of representing the relationship between implicit belief and explicit belief (i.e., ignoring the awareness operator), PWA models resemble the *local reasoning structures* of Fagin and Halpern [16, pp. 58ff], which are intended to represent a 'fragmented' belief system. Indeed, PWA models are essentially generalised local reasoning structures with the addition of awareness functions acting as sieves over the supersets of $\bigcap \mathcal{X}_i(\omega)$. In the same vein, PWA models are also quite similar to the more recent model of implicit and explicit belief developed by Velazquez-Quesada [54]. Velazquez-Quesada's construction works by taking a neighbourhood function $\mathcal{X}_i$ defined for a finite space Ω (interpreted as specifying i's explicit beliefs), and using it to systematically construct another function $\mathcal{X}_i^\star$ which contains Ω and is closed under supersets and binary intersections, interpreted as specifying i's implicit beliefs. In outline, this construction of $\mathcal{X}_i^\star$ is quite similar to my construction of $\|I_i\varphi\|^M$ as a subset of $\{\omega : \bigcap \mathcal{X}_i(\omega) \subseteq \|\varphi\|^M\}$, though there are some important differences. One of these concerns the axiom XI, which is *not* valid on Velazquez-Quesada's class of models. Velazquez-Quesada provides an interesting defence of this characteristic of his model. For present purposes, however, XI seems essential inasmuch as I is understood to represent the informational content contained in an agent's explicit beliefs—every explicit belief that φ contains at least the information that φ, after all.

In terms of the relationship between implicit belief and awareness, the use of the $\mathcal{A}_i$ functions in PWA models is conceptually similar to the way that awareness functions are used in the *Kripke structures for general awareness* of Fagin and Halpern [16]. Fagin and Halpern's models use awareness functions that take sets of sentences as values rather than sets of (possible-worlds) propositions, where those sets of sentences need not be closed under logical equivalence; by virtue of this, they are able to generate extremely fine-grained distinctions between states of awareness. No particular restrictions are placed on Fagin and Halpern's awareness functions, and they define agents' explicit beliefs as a subset of their implicit beliefs—specifically, those the content of which they are aware.

PWA models also share many similarities with the *partitional models* of Fritz and Lederman [20]. (Especially given the presence of *sym* and *con* and with *pla* dropped and stronger rationality conditions placed on the $\mathcal{X}_i$ functions.) Their models are intended to capture a somewhat stronger logic of belief (and knowledge) and its interaction with awareness also within the context of a coarse-grained approach to modelling content. Like Fagin and Halpern, Fritz and Lederman take a notion of implicit belief as a primitive and define explicit beliefs as those states of implicit belief the content of which the agent is aware.

PWA models have much in common with Yalcin's *resolution-sensitive models* of belief [58]. Yalcin uses a partition over logical space to represent different 'resolutions' at which a non-ideal agent might conceive of a space of possibilities. That partition acts to filter between accessible and inaccessible possible-worlds propositions in essentially the same way that the $\mathcal{A}_i$ functions do, or in much the same way that a Boolean sub-algebra of 2^Ω might be used on a probabilistic model of uncertainty to represent the idea that an agent need not have degrees of belief regarding propositions at all levels of specificity. Yalcin is primarily concerned with representing limited informational accessibility under a holistic model of belief for non-ideal agents, where the inaccessibility might result from multiple sources including, for example, attentional or computational limitations, rather than merely conceptual limitations. Many of the motivations for Yalcin's resolution-sensitive

models clearly overlap with those for PWA models, and the former are in effect what PWA models would be if *nax* were removed and the $\mathcal{X}_i$ functions could only ever take singleton sets of propositions as values.

So much for similarities; now let us consider perhaps the main respect in which PWA models differ from a great many alternative models of awareness in the literature. In particular, it is not only common for models of awareness to satisfy at least the axiom DIS_A mentioned earlier, but also more generally the property of *Awareness Generated by Primitive Propositions* (*AGPP*); i.e., that i is aware of every primitive proposition in $\Psi \subseteq \Phi$ if and only if she is aware of every $\varphi \in \mathcal{L}^{AXI}(\Psi)$. (See, for example, [10], [21], [23], [35], and [44].) *AGPP* immediately implies DIS_A, and just like DIS_A it is quite obviously not going to play nicely with any remotely coarse-grained theory of content.

Given what I'm trying to model, however, *AGPP* seems highly implausible. According to *AGPP*, for Iggy to be aware of, say, $A_j p$, he need only be aware of the primitive proposition p. But that cannot be right: he also needs to be aware of the other agent j, and the relation of *awareness*. (Proviso: I am assuming that $A_j p$ is contingent. See Section 6 for discussion.) This suggests two independent problems that we might raise against the principle. First, there are no compelling reasons to think that Iggy must able to represent any agent in **Ag** as an automatic consequence of being able to represent anything at all. Iggy may not even be able to represent *himself*, let alone another agent who he may never have been in any kind of contact with—or who may not even actually exist. Second, Iggy's capacity to entertain p-thoughts doesn't automatically come with the capacity to represent that some agent or other is awareness-related to p, in the specific way that I defined 'awareness' in Section 2.1. (A lot of concepts went into my characterisation of that notion, so an inability to represent the relation of *awareness* wouldn't be at all surprising!) The same, of course, goes for *narrow implicit belief* relation, and even *belief* and *knowledge*. It's not a precondition of agenthood that agents in general should necessarily have the resources to represent those specific kinds of propositional attitude relations.[7] One can readily imagine a member of an alien species, say, or an animal with a less developed theory of mind than our own, who lacks any notion of our specific folk-psychological attitudes and yet can still be said to *have* such attitudes themselves.

Consequently, *AGPP* is too strong as a *general* principle for modelling awareness as I'm understanding it. Likewise, and for the reasons just noted, we will also want to reject each of the following consequences of *AGPP*:

AI_A. $A_i\varphi \to A_i A_j \varphi$
AI_X. $A_i\varphi \to A_i X_j \varphi$
AI_I. $A_i\varphi \to A_i I_j \varphi$

By contrast, SYM and CON_A (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 $\wedge$.

[7] Introspection principles—e.g., $X_i\varphi \to X_i X_i \varphi$ and $\neg X_i\varphi \to X_i \neg X_i \varphi$—would imply that knowledge/belief in any proposition requires higher-order knowledge/belief, and hence awareness, in one's own first-order knowledge/belief states. But these are dubious axioms even for normative logics of knowledge/belief under the assumption of full awareness [56]; they're even less plausible for a descriptively-oriented logic of knowledge/belief that incorporates unawareness such as the one I'm aiming to construct.

4 On the Impossibility of Coarse-Grained 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 awareness and the epistemic attitudes:

PLA. $\neg A_i\varphi \rightarrow (\neg X_i\varphi \wedge \neg X_i\neg X_i\varphi)$
AUR. $A_i\neg A_i\varphi \rightarrow A_i\varphi$
XUI. $\neg X_i\neg A_i\varphi$

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—or even definitional—for any adequate understanding of awareness. See, for instance, [9], [10], [23], [36], [37], [55], though cf. [21], where PLA is derivable 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 an agent i is unaware of φ, then she cannot believe that she doesn't believe φ. AUR (for 'AU Reflection') says that if i is aware of the proposition $\neg A_i\varphi$, then she is *ipso facto* aware of φ itself. The intuitions motivating PLA^b and AUR are 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 intended 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 axioms 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.[8]

In Sections 5–7, we will examine the plausibility of these axioms 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, *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. $X_i\top$
MON. From $\varphi \rightarrow \psi$, infer $X_i\varphi \rightarrow X_i\psi$

[8] 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. For relevant work, see esp. [20], as well as [22], [49], and [55].

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. 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 or MON. Moreover, neither N nor MON (nor PLA^a, for that matter) are essential to generating a serious problem, as the following corollary indicates:

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

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 axioms:

XUI^a. $\neg A_i\varphi \rightarrow \neg X_i\neg A_i\varphi$
XUI^b. $A_i\varphi \rightarrow \neg X_i\neg A_i\varphi$

XUI^a is a simple consequence of AUR and PLA^a, so what XUI actually brings to the table 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, $\vdash_{\Sigma^*} \neg A_i\varphi \rightarrow \neg X_i\top$.*

What's worse, if we add SYM, CON$_A$, and N$_{AX}$ back into the mix, then we again 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$_A$*, and* N$_{AX}$*. Then, for all φ and ψ, $\vdash_{\Sigma^*} \neg A_i\varphi \rightarrow \neg A_i\psi$.*

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

Since PLA^a is secure if anything is, if we want non-trivial awareness then we have a choice:

1. Reject classical logic, and/or
2. Reject at least one of XUI^b, AUR, or PLA^b.

In the following three sections, I will discuss (in turn) XUI^b, AUR, and PLA^b. Overall, I will argue that the possible worlds theorist is entitled to reject at least one of XUI^b, AUR, or PLA^b, and perhaps all three. However, matters are complicated: which of those three axioms we ought to reject in particular will depend partly on (i) how we want to interpret X, (ii) the specific coarse-grained theory of content that we want to adopt, and (iii) whether there exist any necessarily unentertainable propositions.

5 Critical Discussion: XUI^b

Assume first that we want X to specifically 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 axiom, there is no plausible way out of the triviality result via XUI^b under such an interpretation.

On the other hand, assume that we want X to represent belief. Under this interpretation, XUI^b says that if i is aware of φ, then she does not explicitly believe that she's not aware of φ. And this does not seem especially plausible.[9]

It would be *odd*, perhaps, for Iggy to have false beliefs about his own state of awareness. After all, if Iggy is aware of φ, then you might think that he has access evidence justifying the belief that he's aware of φ—at least where φ is a proposition that Iggy can actually entertain. (Recall from Section 2.1 that entertainability in principle doesn't mean entertainability in practice, and Iggy can of course only believe what he can in fact entertain.) But XUI^b says that Iggy *cannot* falsely believe that he is in some state of unawareness. And even if he does have evidence that he's aware of φ, he might have even more evidence that he's *not* aware of φ. Or perhaps he has evidence that he's aware of φ, but he doesn't weigh up that evidence appropriately. As a non-ideal agent, then, Iggy can presumably have all sorts of absurd and unjustified beliefs that fly in the face of his evidence. Why should his beliefs about awareness be any different?

Consider, for instance, the following case:

> **Example 1.** After long discussions on the matter with leading philosophers and psychologists, Iggy 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 inaccurate theory of how our minds work.[10] Iggy has thought a lot about folk psychology and about his stance on it, so he's clearly aware of the proposition *Iggy does not believe that folk psychology is an accurate theory*. However, he doesn't believe that he's aware of that proposition—in fact, he strongly (and wrongly) believes that he isn't.

Let φ be the proposition that *Iggy does not believe that folk psychology is an accurate theory*. We might view this as a case where Iggy has overwhelming evidence that $\neg A_i\varphi$, or one where he has insufficient evidence that $\neg A_i\varphi$ but believes it

[9] 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., [10, p. 3], [42], [49, p. 516] and the models outlined in [44].

[10] 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., $\neg A_i\varphi$, $A_i\varphi \to A_i\varphi$, or $A_i\varphi \vee \psi$ (i.e., if ψ is true). 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.

anyway. Either way, we have a plausible situation where $A_i\varphi \wedge X_i\neg A_i\varphi$ is true, and thus a counterexample to XUI^b.

So, here is our first lesson: while XUI^b is clearly 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. Note that the point being made here is entirely independent of how fine-grained or coarse-grained we take mental contents to be. XUI^b is dubious for non-veridical belief given any position on the granularity of mental content, and should therefore be dropped regardless of whether we want to stick with a classical logic or not.

6 Critical Discussion: AUR

Since AUR doesn't interact in any interesting way with T_X, we can consider its plausibility independently of how we interpret the X and I operators. Nevertheless, AUR turns out to be much trickier than XUI^b to evaluate, and doing so will involve opening more than one can of worms. The large number of issues here necessitates some brevity, and I will not pretend to have stated the last word on any of them.

It's unsurprising that AUR is plausible given a fine-grained theory of content, so the point of the present section is only to consider whether and to what extent AUR can also be considered plausible from the perspective of a coarse-grained theory of content. That is, the question I'll be considering is: *supposing we've already accepted a coarse-grained theory of content, should we also accept* AUR? By the end of the section I will conclude that (i) counterexamples to AUR seem to exist on a 'metaphysical' conception of coarse-grained content, while (ii) on an alternative 'epistemic' conception of coarse-grained content, AUR is defensible.

6.1 Metaphysical conceptions of coarse-grained 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. I suspect that most readers, when they think of a coarse-grained theory of content, will have this kind of conception in mind; it is the kind of account most typically (but not necessarily) associated with externalism about content.[11]

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. Such counterexamples can exist only where the ψ in question is either metaphysically contingent or impossible: if ψ and $\psi \leftrightarrow A_i\varphi$ are both metaphysically necessary, then so is $A_i\neg A_i\varphi \rightarrow A_i\varphi$. I will discuss whether there are any necessarily unentertainable propositions separately (Section 6.2), so here we will focus only on those cases where the ψ in question is contingent.

[11] Thanks to [*anonymised*] 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.

Prima facie, counterexamples of this form should be easy to find: for any given proposition ψ, there will usually be many ways—different modes of presentation—under which a different agents might entertain something metaphysically equivalent to ψ despite having very different conceptual resources. I think that the relevant counterexamples *probably* do exist. But their existence is not quite as easy to establish as first appearances suggest.

Consider the following case:

> **Example 2.** Iggy has just made first contact with an alien from the planet Gzorp, with whom he's attempting to communicate. Gzorp is 65.24 light years away from Earth, bearing 92.45° at an elevation of 30.13°. In its own language, the Gzorpian asserts that *Iggy is not aware of the proposition that skirnobs are poisonous*, where a *skirnob* is a kind of Gzorpian fruit found all over the planet. Iggy does not understand, of course, but he knows that the Gzorpian just asserted something, and that whatever it is, it's true.

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 $A_i\psi$, and $\psi \leftrightarrow \neg A_i\varphi$. If it's also the case that $\neg A_i\varphi$, then we have a counterexample to AUR.

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

1. The kind of fruit the Gzorpian just referred to is poisonous.
2. The kind of fruit with purple stripes and pink feathers is poisonous.
3. The kind of fruit, an instance of which is 65.24 light years away from me bearing 92.45° at an elevation of 30.13°, is poisonous.
4. The kinds of things referred to by the sound /skəːnɒb/ in the language of *those* things are poisonous.
5. The kinds of things in the category my community refers to when they make the sound /fruːt/, an instance of which is 6.17218×10^{17} metres away from me bearing 92.45° at an elevation of 30.13°, are poisonous.

Generally: give Iggy some basic 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 Iggy won't in principle have the resources to think about under some mode of presentation or other (cf. [26]). He almost certainly never *will* have all these thoughts, but he *could*—and that's all that's required for awareness. So, perhaps Iggy *could* be aware of the proposition *Iggy is not aware of the proposition that skirnobs are poisonous* without being aware of *skirnobs are poisonous*, but this would require that he's quite representationally impoverished indeed.

I suspect that it is *possible* for agents to be impoverished in this way, but we couldn't expect it of any (or many) *actual* human beings. Thus, while we have here a potential counterexample to AUR, the case also points to a kind of practical triviality for the notion of awareness under this conception of coarse-grained content. That is, awareness can come *very* cheap on the metaphysical conception, and this cheapness serves to undermine the theoretical importance of the notion. Even deep in the jungle, Iggy could still think a thought with the same

content as that which we would ordinarily express upon uttering 'The habitat of the eastern grey kangaroo extends as far north as the Cape York Peninsula'; in fact there's dozens of obvious and easy ways by which he could do this, though Iggy himself wouldn't recognise them as such.

The problem, of course, is that the flexibility in constructing alternative modes of presentation by which one might entertain $A_i\varphi$ will tend to go hand-in-hand with the same degree of flexibility in constructing alternative modes of presentation for entertaining φ itself. It's easy to imagine various ways of describing the set of metaphysically possible worlds where $A_i\varphi$ is true that don't specifically mention φ; what's a lot less easy is to imagine an agent with the capacity to entertain the former without also having *any* capacity to entertain something metaphysically equivalent to the latter.

One more example to drive the point home:

> **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. Iggy is a brilliant neuroscientist, and can represent himself as being in state S.

This time, letting ψ denote that set of worlds where Iggy is in state S, $\psi \leftrightarrow A_i\varphi$ is necessary. And since $A_i\psi$, so we also have $A_i\neg\psi$. But do we have $A_i\varphi$? This would be easier to argue for if we were internalists about content. In that case S could be identified with some internal (presumably neurological) state, and it's quite plausible that Iggy 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 *inside my head is like that* would suffice.) But internalism sits poorly with the metaphysical conception, and few adherents to the latter would want to commit themselves to the former. Externalists will have a much tougher time of showing that $A_i\varphi$ holds in this example. After all, on an externalist's theory, 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. And being able to represent *that* kind of state without being able to represent *skirnobs* in one way or another is not obviously possible.

There are additional examples we could consider, involving for example Burgean social externalism and the apparent capacity to entertain a thought about *arthritis* without really knowing what that is merely by being situated in the relevant linguistic community. I will not try to consider them all. What is clear is that there is an argument—or rather, several arguments—to be made for the existence of counterexamples to AUR under the metaphysical conception of coarse-grained content. With that said, there is also a genuine concern that awareness is a little *too* cheap on the metaphysical conception, at least for ordinary agents whom we can expect to have reasonably rich basic representational resources.

6.2 Necessarily unentertainable propositions

Perhaps there is an easier route to finding counterexamples to AUR: if, for some agent i, there are any propositions φ of which i is *necessarily* unaware, then $\neg A_i\varphi$ is necessary, and there will be (vastly many) ways that i might entertain $\neg A_i\varphi$ without entertaining φ. (For example, i could entertain $\neg A_i\varphi$ by entertaining $\psi \vee \neg\psi$,

for any ψ that i can in fact entertain.) But are there any propositions φ and agents i such that it's *metaphysically impossible* that i lacks the basic representational resources to entertain φ? It's not obvious that there are.[12] Let me consider two general kinds of argument for necessarily unentertainable propositions.

First, you might think that some propositions cannot be entertained by beings us as a result of our cognitive limitations. For instance, due to limited resources of memory and processing, there may be some φ which are simply too complex or specific for an ordinary agent like Iggy to entertain. Of course, limitations of memory or processing are insufficient to make the point: these limitations are contingent, and there are no good reasons to suspect that there aren't any possible worlds where they have been overcome (including, potentially, in our own future). And as we've already discussed, the notion of awareness that we're playing with here isn't bound by the contingent limitations of the ordinary human brain (Section 2.1). But there are other examples in the vicinity that might deserve some more weight. In particular, you might think that there are hard-wired limitations on the kinds of phenomenological experiences we might undergo [39, pp. 90ff]. Given the kind of being that Iggy is, perhaps he cannot—as a matter of metaphysical necessity—know what it is like to be a bat. Now, *if* (i) it is metaphysically necessary of Iggy that he is a human, (ii) it is metaphysically essential to being human that one's brain is hard-wired in *this* kind of way (i.e., such as to preclude the possibility of undergoing certain kinds of experiences), and (iii) 'what it is like to be a bat' can be cashed out in terms of some propositional content (or contents), *then* we might have a counterexample to AUR here. I will leave this matter hanging, though it should go without saying that not one of (i)–(iii) is unassailable.

The second kind of argument purports to establish that there aren't enough entertainable propositions to go around. Perhaps the best known of these can be found in Lewis [33, pp. 104–7], as part of his response to the Russell-Kaplan paradox (cf. [13], [29]). Lewis gives three reasons for thinking that there are unentertainable propositions. The first is that it provides a way out of the paradox, though few commentators have been convinced by this—at most it helps with a very specific version of the paradox, but doesn't address the deeper problem (e.g., [3], [7], [53]). The second reason rests heavily on the specifics of Lewis' views on mental content and the role that relative naturalness plays within it. I'll describe this briefly below, but a full discussion would take us too far afield. The third and final reason arises from Lewis' functionalism more generally, and it's this reason that we'll focus on here.

As Lewis points out, if one accepts a certain kind of functionalism about what it is to believe (or desire, be uncertain about, etc.) a proposition—I'll say more about the functionalist assumptions in a moment—then there can be no more states of *believing* (or *desiring*, *being uncertain about*, etc.) a proposition φ than there are functional roles with which to define those states. However, while there are probably at least $\beth_3$ sets of possible worlds, Lewis also asserts that there are probably no more than countably many functional roles. So, there are more

[12] For a recent and thorough treatment of several arguments relating to the present discussion, see [26]. It's important to note, however, 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—i.e., whether we might express or entertain every proposition *given the way we actually are*. His concern is not directly on whether there exists any propositions that are necessarily unentertainable *tout court*, or necessarily unentertainable for some agent i.

possible worlds propositions than there are distinct states of *believing that* φ. Most propositions are in fact unbelievable. Likewise, most are undesirable, and most we cannot have any uncertainty about, and so on. If we add now the reasonable premise that if it's possible for i to be aware of φ, then it's possible for i to believe (or desire, be uncertain about, etc.) that φ, then we get the conclusion that there are propositions of which i cannot possibly be aware.[13]

There are two obvious ways to resist this conclusion. One would be to deny that there are fewer functional roles than there are propositions. Lewis does not say much to support this, only that he 'cannot see the slightest *prima facie* reason to think that there are even uncountably many of the definitive roles' (pp. 106–7). The thought, perhaps, was this: for the typical functionalist of the reductive physicalist variety, a definitive functional role for a given mental state S is characterised in terms of causal relations between that state and some range of possible sensory inputs, behavioural outputs, and other (physical, e.g., neurological) states. Suppose therefore that there are only finitely many relevantly distinct possible sensory inputs, behavioural outputs, and other neurological states. This is plausible enough for beings *like us*; we have limited capacities for making meaningful distinctions between sensory inputs, for example, and our behavioural dispositions can only get so fine-grained. If so, then for each finite n there would be no more than finitely many n-ary relevantly distinct relations that hold between those inputs, outputs, and other states; and hence, no more than countably many functional roles could be characterised in terms of those relations (or continuum many if we allow also relations which take a countably infinite number of arguments).

The problem, again, is that it's not clear whether these finite limitations are metaphysically essential to thinkers in general. As noted in ([7], pp. 91–2) if there can be arbitrarily complex infinitary beings, then there's no reason why there cannot be arbitrarily many functional roles. If there are k relevantly distinct input, output, or neurological states, then there are (at least) 2^k n-ary relations between those states.[14] And to reiterate the point made earlier, it's at least arguable that Iggy, or any other agent i, *could* have been such an infinitary being, capable of making arbitrarily fine-grained distinctions in perception or behaviour, or with an infinitely complex neurological state. Alternatively, given a non-reductive functionalism—such as that found in Schwitzgebel's dispositionalism for instance; see [46]—we might allow definitive functional relations to hold between a richer variety of phenomena. Schwitzgebel himself characterises 'beliefs' partly in terms of their relationships with non-reduced phenomenological properties, and (for all we know) there may well be an arbitrarily large number of these.

A second way to resist Lewis' conclusion would be to deny the implicit assumption that for each φ such that an agent can believe (or desire, be uncertain about, etc.) φ, there must be a separate functional role by which *believing* (or *desiring*, or *being uncertain about*, etc.) φ is to be defined. There are multiple ways you might go about this, but I'll just discuss the one that I find the most interesting. In particular, you might think that instead of each individual attitude

[13] Lewis conducts his discussion in terms of 'thoughts' and 'thinking' rather than 'beliefs', 'desires' and 'uncertainties', as I've done here. This won't make any difference to my response, but it will help to avoid some ambiguities.

[14] Assume, e.g., that there is at least one distinct binary relation on any set $\mathcal{S}$ for every non-empty set of ordered pairs of members of $\mathcal{S}$; then there are at least as many binary relations on $\mathcal{S}$ as there are non-empty subsets of $\mathcal{S} \times \mathcal{S}$.

receiving a separate definition, it's entire or 'holistic' belief-plus-desire states that are functionally defined.

Lewis himself seems to have held a view like this. Roughly, given a sequence of perceptual or evidential inputs E and a suite of appropriately characterised behavioural dispositions D, we are to assign to an agent (or an appropriate state of the agent, most likely a brain state; see [32], p. 373) the maximally 'eligible' set of credences and utilities that rationalises D given E; see [30], [32]. Lewis thought that, given *any* D and E pair, there will always be multiple sets of credences and utilities that rationalise D given E. Hence, he thought that there are more holistic credence-utility states than there are definitive functional roles, and since only one or at most a relative few of those will be maximally 'eligible', there will thus be some credence-utility states which are never assigned at any possible world. I have argued elsewhere that this was a mistake (see [*anonymised*]; cf. also [45]), but that's neither here nor there. Even supposing it's true that there are, let's imagine, only countably many definitive functional roles with which to define the possible assignments of holistic credence-utility states, it simply does not follow from this that there are only countably many propositions to which an agent might have credences and/or utilities. Similarly, imagine that there are only countably many pairs of sets of beliefs and desires that can be functionally characterised—this is entirely consistent with saying that there are an arbitrarily large number of distinct propositions which might belong to one of those sets.

Again, there is more that could be said on each of the above points. Perhaps there are brute necessities, for instance, and one of those brute necessities is that i never happens to be aware of the proposition φ. I don't see any way that we could rule that out, short of settling once and for all the difficult matter of whether there are any brute necessities *tout court*. Or perhaps there are facts about transworld identity which preclude any human agent from possibly being like a bat, or from having an infinitely complex brain. I am not convinced. More importantly, it's not *obvious* that there are any necessarily unentertainable propositions. The easy route to finding counterexamples to AUR doesn't seem so easy after all.

6.3 Epistemic conceptions of coarse-grained content

So much for the metaphysical conception. Let's now consider an alternative, *epistemic* conception of coarse-grained content.[15] 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 modes of presentation are metaphysi-

[15] I include here only a bare-bones development of the epistemic conception. More detailed developments can be found in [5], [6], [7], [27], [28]. I have defended the epistemic conception elsewhere; see [*anonymised*].

cally equivalent, yet the equivalence of those modes of presentation 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. (Though this is not true of everyone; cf. [41].) 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 theory to content, let me note some points before moving on.

First, one should not assume that the epistemic conception involves merely supplementing the original space of metaphysically possible worlds with a number of metaphysically impossible worlds so as to let us 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 the thought that *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 is not merely the metaphysical conception with a few extra impossible worlds.

Indeed, on so-called 'one-spaceist' views (see e.g., Jackson in [28] and Chalmers in [6, 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 is more (or less) fine-grained than the metaphysical conception—contents on either conception are simply subsets of a single space of worlds, Ω, so each posits a notion of content which is exactly as coarse-grained as that which is posited by the other.

Supposing then that classical logic is a priori, if we adopt the epistemic conception then 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 earlier list of modes of presentation for thinking that *skirnobs are poisonous* are a priori equivalent, so each will hold relative to a different subset of Ω. Moreover, cases like Examples 2 and 3 fail to generate counterexamples to AUR, for the proposition ψ that Iggy 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 *Iggy is in brain state S* if and only if *Iggy is aware of the proposition that Iggy is aware of the proposition that skirnobs are poisonous*. Awareness is not as cheap on the epistemic conception as it is on the metaphysical 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.[16]

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 Iggy entertain some content ψ that has the very same cognitive significance as *Iggy is awareness-related to* φ, without having the conceptual resources to represent *Iggy*, *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 Iggy is aware of φ under a mode of presentation that's a priori equivalent to *Iggy 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 on the epistemic conception that there are no φ such that it's a priori that Iggy is not aware of φ. Arguments from cognitive limitations don't seem to get any grip: for any limitations that are supposedly essential Iggy *qua* human being, it is not *a priori* for Iggy that he is subject to those limitations. For all he knows a priori, he *could* have been a bat. (See also [7], §9, for a detailed discussion of whether there are any a priori unentertainable propositions.)

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

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

Assuming that Iggy's designation succeeds—i.e., there is indeed a unique φ that satisfies the description—then Iggy'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 Iggy that he is not aware of φ, because it is not a priori what proposition 'Silly' picks out (if in fact it picks out any proposition at all). What Iggy knows a priori is that if 'Silly' picks out something, then whatever proposition it picks out, he is not aware of that proposition. At one epistemic possibility ω_1, 'Silly' might designate φ_1; at ω_2, φ_2. It's not a priori what proposition 'Silly' picks out, so there's no specific φ that Iggy knows he's not aware of, and AUR is safe.

7 Critical Discussion: PLA^b

Finally, let us consider PLA^b. As with Section 6, I will discuss the plausibility of this axiom from the perspective of one who has already accepted a coarse-grained theory of content. There are two main cases: those where there are counterexamples to AUR (e.g., if we've adopted a metaphysical conception), and those where there are no counterexamples to AUR (e.g., if we've adopted an epistemic conception).

[16] 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. [5, §9].

7.1 Counterexamples from AUR to PLA^b

I will here argue that any counterexamples to AUR can be expected to generate a counterexamples to PLA^b, at least inasmuch as X represents knowledge.

This should be unsurprising: the potential counterexamples to AUR that were discussed in the previous section can quite clearly be tweaked to provide counterexamples to PLA^b. Consider Example 2, and substitute '*Iggy is not aware of the proposition that...*' for '*Iggy does not know that...*'; everything else about can be left unchanged. The case would be one in which $X_i\psi$ is true and $\psi \leftrightarrow \neg X_i\varphi$ is necessarily true, which combined with $\neg A_i\varphi$ would constitute a counterexample to PLA^b. (The substitution of 'aware of' for 'knows that' should make no difference to the plausibility of $\neg A_i\varphi$ in that example.) Similarly, if it turns out that there is some φ of which a given agent i is necessarily unaware, then φ is *ipso facto* also something that i necessarily cannot know or believe. In that case, $\top \leftrightarrow \neg X_i\varphi$ is necessary, leading directly to a counterexample against PLA^b.

However, there are also more general reasons to think that there are no counterexamples to AUR that do not also generate counterexamples to PLA^b, in the presence of T_X. Thus, at least where X represents knowledge, we should be willing to reject AUR only if we're also willing to reject PLA^b. (This point will not be especially important for the rest of my discussion, so if you're satisfied with what's been said already then you may wish to skip directly on to Section 7.2.)

First of all, I assume that if $X_i\neg A_i\varphi \wedge \neg A_i\varphi$ can be true, then $X_i\neg X_i\varphi \wedge \neg A_i\varphi$ can also be true. If Iggy can know that Iggy isn't aware of φ (despite not being aware of φ), then Iggy can know that Iggy doesn't know that φ (despite not being aware of φ). After all, the latter is an obvious consequence of the former, and Iggy should be able to know any obvious consequences of what he knows if he's also aware of all the relevant propositions. And if being unaware of φ doesn't prevent him from being aware of $\neg A_i\varphi$, then there's no reason to think that it should prevent him from being aware of $\neg X_i\varphi$.

Second, note that we have a counterexample to AUR only if there is some possible situation where $A_i\neg A_i\varphi \wedge \neg A_i\varphi$; and a counterexample to PLA^b only if there is a situation where $X_i\neg X_i\varphi \wedge \neg A_i\varphi$. So let $\psi = \neg A_i\varphi$, and assume that there does exist some counterexample to AUR while there are also no counterexamples to PLA^b. Obviously this requires that ψ be contingent, given the points made above; likewise φ must also be contingent, else $\neg A_i\varphi$ would be incompatible with $A_i\neg A_i\varphi$. In light of the assumption of the previous paragraph and the fact that $X_i\psi$ already implies ψ, this would have to mean that $A_i\psi \wedge \psi$ somehow precludes the possibility of $X_i\psi$—which would be very strange, since if ψ is true *and* i is aware of ψ, then i should be able to know ψ. It would be one thing if the truth of ψ for some reason precluded awareness of ψ, which would in turn naturally rule out any knowledge of ψ; but if the truth of ψ is compatible with awareness of ψ, then it should also be compatible with knowledge that ψ. (Simply imagine a case where $A_i\psi \wedge \psi$, and i has, let's say, plenty of reliable testimonial evidence that ψ.)

Furthermore, since $\neg A_i\psi$ already implies $\neg X_i\psi$, the assumption that AUR faces counterexamples while PLA^b doesn't would mean that ψ alone implies $\neg X_i\psi$; so, $X_i\psi$ would have to be impossible *tout court*. By PLA^b, $\neg A_i\psi$ implies $\neg X_i\neg X_i\psi$, so $\neg A_i\psi$ implies $\neg X_i\top$; and by N_{AX}, $\neg X_i\top$ implies $\neg A_i\top$. Putting these two points together, $\neg A_i\psi$ would imply $\neg A_i\top$; and we already know that $A_i\psi$ implies $A_i\top$. Hence, $A_i\top \leftrightarrow A_i\psi$ would be necessary. Yet ψ cannot be impossible, by

hypothesis; likewise, ψ cannot be necessary, since then $X_i\top$ would be impossible and so would $A_i\psi$, which again it isn't by hypothesis. Thus, for PLA^b to avoid falling foul of the whatever counterexamples exist for AUR, we would need that there are no propositions φ of which i is necessarily unaware, and for each and every φ such that $A_i\neg A_i\varphi \wedge \neg A_i\varphi$ is possible, $A_iA_i\varphi \leftrightarrow A_i\top$ is necessary.

In sum: for there to be counterexamples to AUR yet no counterexamples to PLA^b, there would need to be some pair of contingent propositions, φ and ψ, such that (i) ψ is true just in case i is not aware of φ, (ii) i is aware of ψ whenever i is aware of anything whatsoever, and (iii) i cannot know ψ even when ψ is true and i is aware of it. I think we can safely assume that no such pair φ and ψ exists, even setting aside the already very strange property (iii). For what reason do we have to think that i's being aware of anything at all would require being aware of $\neg A_i\varphi$, for some contingent φ? The only somewhat plausible suggestion for what φ could be in this case is $A_i\top$, or (what amounts to the same thing given SYM and CON$_A$) '*i is aware of something*'. But I have argued already in Section 3.3 that i's being aware of something or other doesn't presuppose the capacity to represent either i or *awareness* itself, so we don't get $A_i\neg A_iA_i\top$ as a consequence for i's being aware of anything at all out of some simple iterative principles like AI$_A$. And if I am right about that, then more generally we shouldn't expect awareness of any contingent proposition $\neg A_i\varphi$ to be a consequence of awareness *simpliciter*—after all, any other way of entertaining that same proposition would require some other suite of representational resources which i may well also lack.

7.2 The implausibility of 'Plausibility'

So the possible worlds theorist should reject PLA^b, if they think that there are in fact some counterexamples to AUR. However, I have also argued that, at least on an epistemic conception of coarse-grained content, AUR is plausible. Consequently, I will here argue that where X represents knowledge, if the possible worlds theorist is willing to accept AUR, then they should reject PLA^b—indeed, in part because of the acceptance of AUR.

This might seem a little surprising. AUR and PLA^b rest on more or less the same intuition. So, just as you might expect that if there are any problems for the former then there will be analogous problems for the latter, likewise you might expect that if there are any problems for the latter then there will be analogous problems for the former. And if we've decided that AUR is plausible after all, then shouldn't we do the same for PLA^b?

The problem with this thought is that AUR and PLA^b are not entirely symmetrical in all relevant respects, *once* T_X *is factored into the equation.* In particular, AUR implies the existence, for each agent i, of a specific class of propositions that i necessarily cannot know. Assume that there are no counterexamples to AUR. Then, for any φ, suppose that i knows that i is not aware of φ. Knowledge presupposes awareness, so i must be aware of i's being unaware of φ, and therefore he must be aware of φ. However, by veridicality, i cannot be aware of φ. Contradiction. Consequently, *if* AUR is accepted, then there are certain propositions that i cannot know—i.e., anything necessarily equivalent to $\neg A_i\varphi$—simply because the truth of those propositions is incompatible with i's knowledge thereof.

On the other hand, PLA^b does not imply the existence of any propositions of which i is necessarily unaware. And this is an important asymmetry. For we have seen that AUR is plausible from the perspective of a coarse-grained theory of content only if there are no propositions φ of which i is necessarily unaware—simply because, if $\neg A_i\varphi$ is impossible, then i might come to be aware of $\neg A_i\varphi$ in a myriad of ways without being aware of φ (or *himself*, or the *awareness* relation). By the same token, PLA^b is plausible from the perspective of a coarse-grained theory of content only if there are no propositions that i cannot possibly know—simply because, if $X_i\varphi$ is impossible, then i might come to know $\neg X_i\varphi$ in a myriad of ways without being aware of φ (or *himself*, or the *awareness* relation). Thus, on a coarse-grained theory of content, the acceptance of AUR generates problems for PLA^b, but not *vice versa*.

But we don't need to go via AUR to find problems with PLA^b. Given T_X, there are plenty of other kinds of impossible knowledge states. Suppose we introduced into the language e_i, interpreted as 'i exists', and stipulate that it's true at a world ω just in case $\mathcal{A}_i(\omega)$ is non-empty. (An agent who's not aware of anything whatsoever is not an agent at all, so the agent i exists only if i is aware of something.) Now consider $X_i\neg X_i\neg e_i$: does $A_i\neg e_i$ follow? No: the agent i might come to know $\top$ in any number of ways, and *ipso facto* come to know $\neg X_i\neg e_i$, without necessarily being able to represent himself or his own (in)existence. Or suppose that b is true just in case *there are beliefs*; then $X_i\neg b$ is impossible, $T \leftrightarrow \neg X_i\neg b$ necessary, yet knowing $\top$ doesn't imply being aware of b: having beliefs does not require being able to think thoughts about the existence of beliefs.

8 In Defence of Classical Logic

The points made in Section 6 and Section 7 obviously hang on the prior acceptance of a coarse-grained theory of content in some form or another. I don't expect to have convinced anyone who thinks that the contents of thought cut finer than necessary equivalence that they ought to reject AUR and/or PLA^b. To such a reader it will likely appear that I am seeing counterexamples to AUR and PLA^b where I really ought to be seeing counterexamples to the coarse-grained theory of content. I've argued, for example, that since $\neg X_i\neg A_i\varphi$ is necessary and contents are coarse-grained, therefore i can know $\neg X_i\neg A_i\varphi$ without being aware of $A_i\varphi$. But perhaps I should have argued instead that since i can't know and be aware of $\neg X_i\neg A_i\varphi$ without being aware of $A_i\varphi$, therefore contents are not coarse-grained. (This *was* the essential point of Dekel *et al.*'s argument, after all.)

But therein lies I think the most important lesson of the foregoing discussion: to the extent that incorporating some representation of unawareness into classical logics of belief and knowledge presents possible worlds theorists with any problems at all, it doesn't seem to add any *specific* issues over and above the more general concerns about hyperintensionality of which we are all already aware. For the possible worlds theorist, who has considered the arguments and intuitions in favour of fine-grained contents and found them wanting, it's not at all troubling to be told that there's something counterintuitive to saying that i can know and be aware of $\neg X_i\neg A_i\varphi$ even while unaware of $A_i\varphi$, given that $X_i\neg A_i\varphi$ is impossible. This is precisely on a par with being told that one can know $p \vee \neg p$ even while one is

unaware of p—and no possible worlds theorist is going to give up on their position because of that! Or consider again the axiom DIS_A:

DIS_A. $\quad A_i(\varphi \wedge \psi) \rightarrow (A_i\varphi \wedge A_i\psi)$

There is no denying that DIS_A has *prima facie* plausibility. But that plausibility stems entirely from the intuition that mental contents cut finer than necessary equivalence—that to be aware of $\varphi \wedge \psi$ is to be aware of some structured entity which has φ and ψ as its essential parts. Any triviality argument against a coarse-grained account of unawareness that's based on DIS_A (like the one we saw in Section 3.1) is by virtue of this no more compelling a reason to reject the coarse-grained view than the already apparent strangeness of the claim that by knowing $p \vee \neg p$ you therefore also automatically know $\neg(q \wedge \neg q)$.

If there were something *special* about the phenomenon of unawareness itself which rendered it incompatible with classical logic—something separable from presuppositions about granularity—then that would be a major blow for the possible worlds theorist. But there is nothing *new* here, just an old problem dressed up in new clothes. And, importantly, possible worlds theorists have a suite of tools to help address the problems that arise from the apparent hyperintensionality of thought. Given a Two-Dimensionalist approach to mental content and an appropriately characterised space of epistemically possible worlds, for example, many of the classic 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 propositional contents are sets of possible worlds. And conversational pragmatics can help deal with linguistic intuitions about the substitutability of that-clauses within the context of propositional attitude verbs even in cases of logical or a priori equivalence (see esp. [50], [51]).[17]

The real concern for coarse-grained theories of content arises when they are taken in conjunction with a plausible theory of action; *viz.*, that agents will typically act so as to maximise their desire or preference satisfaction given the way they believe the world to be (cf. [52]). If a theory of content has empirically false behavioural implications then it must be rejected—and, *prima facie*, the typical subject doesn't seem to act as we might expect given beliefs in every necessary truth unless we posit that they have very strange 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 doesn'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 the necessary truth—the task is to attribute some plausible set of beliefs and desires that rationalises the time spent discovering yet another complicated way to say $\top$.)

For AUR and/or PLA^b to generate empirical problems for the coarse-grained theory along similar lines, 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

[17] 'Fragmentation' also helps deal with additional concerns relating to information access and logical closure properties. However, these are much less pressing issues for classical logics *in general*, which are only committed to closure under necessary equivalence. Similarly for PWA models, where explicit beliefs are not closed under implication and implicit beliefs are only closed under implication relative to awareness.

for i to be aware of φ (or such that i cannot know that she is unaware of φ), tends to behave as if she doesn't believe that she's unaware of φ (or as if she doesn't believe that she doesn't know she's unaware of φ). But how 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 would be generated by those 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 state of awareness in this respect. More generally, according to the view of rational action that is supposed to generate the problem, agents' actions are a response to those things they believe they can change so as improve their situation. So considered 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 $\neg A_i\varphi$, or $\neg X_i \neg A_i\varphi$, 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 so, we'll need to either adopt a more fine-grained approach to content, or revise our theory of action. I take this is still an open empirical question. The success of standard models of decision-making—which generally make use of coarse-grained content—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. And whatever issues such models may or may not have, they are independent of considerations arising from unawareness.

9 A Model of Rational Belief with Non-Trivial Unawareness

Let's take stock. Where X represents knowledge, XUI^b is plausible but PLA^b isn't. Where X represents belief, XUI^b is implausible, while PLA^b seems about as plausible as AUR is. Whether AUR is plausible is independent of how we interpret X, but does seem to depend on (a) the particular theory of coarse-grained content, and (b) whether there are any necessarily or a priori unentertainable propositions. Thus, where X represents belief, the status of AUR and PLA^b is uncertain.

In this final section, I will prove that supplementing Σ with AUR and PLA^b is consistent with non-trivial awareness in the absence of XUI^b. In fact, I'll prove something stronger than that—suppose we add the following three axioms to Σ:

AR_A. $A_iA_j\varphi \rightarrow A_i\varphi$
AR_X. $A_iX_j\varphi \rightarrow A_i\varphi$
AR_I. $A_iI_j\varphi \rightarrow A_i\varphi$

Given SYM, AR_A implies AUR; and given SYM and PLA^a, AR_X gets us to PLA^b. The three in combination say that for all agents i and j, if i is aware of j's having some attitude (A, X, or I) regarding φ, then i must herself be aware of φ. I take it that this is a natural generalisation of the shared intuition that motivates both

PLA^b and AUR—so, if we can show that the combination of AR_A, AR_X, and AR_I is compatible with non-trivial awareness in the absence of XUI^b, then we'll also have given strong reasons to believe that there are no other triviality results in the vicinity of Dekel *et al.*'s that rest on the same intuitions but don't go through XUI^b.

Let $\Sigma^\dagger$ refer to the system $\Sigma \cup \{\mathrm{AR}_A, \mathrm{AR}_X, \mathrm{AR}_I\}$, and let $\boldsymbol{\mathcal{M}}^\dagger$ refer to that class of PWA models which satisfies the following additional constraints:

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

The following is then easy to prove given the proof of Theorem 1:

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

Proof See Appendix B.

We can now show that there are PWA models of (rational) believers belonging to $\boldsymbol{\mathcal{M}}^\dagger$ which allow for non-trivial awareness. For the model $M^e = (\Omega, \mathcal{X}_i, \mathcal{A}_i, \pi)$, let $\Omega = \{\omega_1, \omega_2, \omega_3\}$, and $\Phi = \{p, q, r\}$. We will suppose that $\pi(p) = \{\omega_1\}$, $\pi(q) = \{\omega_2\}$, and $\pi(r) = \{\omega_3\}$, and we let Iggy's awareness and belief functions be defined as follows:

$$\mathcal{A}_i(\omega_n) = \begin{cases} \{\emptyset, \{\omega_1\}, \{\omega_2, \omega_3\}, \Omega\}, & \text{if } n = 1 \\ \{\emptyset, \Omega\}, & \text{if } n = 2 \\ 2^\Omega, & \text{if } n = 3 \end{cases}$$

$$\mathcal{X}_i(\omega_n) = \begin{cases} \{\{\omega_2, \omega_3\}, \Omega\}, & \text{if } n = 1 \\ \{\Omega\}, & \text{if } n = 2 \\ \{\Omega\}, & \text{if } n = 3 \end{cases}$$

The world of interest is ω_1, where Iggy's awareness is non-trivial: he is able to draw a distinction between p-worlds and $\neg p$-worlds, but he is unable to draw a distinction between q-worlds and r-worlds.

From the model M^e, Theorem 4 follows:

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

Proof The model obviously satisfies *pla*, *sym*, *con*, and *nax*. That M^e satisfies *ara* can be seen by noting first of all that:

$$\{\omega : P \in \mathcal{A}_i(\omega)\} = \begin{cases} \Omega, & \text{if } P \in \{\emptyset, \Omega\} \\ \{\omega_1, \omega_3\}, & \text{if } P \in \{\{\omega_1\}, \{\omega_2, \omega_3\}\} \\ \{\omega_3\}, & \text{otherwise} \end{cases}$$

So there are exactly three propositions P (i.e., Ω, $\{\omega_1, \omega_3\}$, $\{\omega_3\}$) such that for some φ, $P = \|A_i\varphi\|^{M^e}$. In all three cases it's easy to check that if Ω, $\{\omega_1, \omega_3\}$ or $\{\omega_3\}$ belongs to $\mathcal{A}_i(\omega)$, then the required proposition P also belongs to $\mathcal{A}_i(\omega)$. Similarly for *arx*:

$$\{\omega' : P \in \mathcal{X}_i(\omega)\} = \begin{cases} \Omega, & \text{if } P = \Omega \\ \{\omega_1\}, & \text{if } P = \{\omega_2, \omega_3\} \end{cases}$$

So there are exactly two propositions P (i.e., Ω, $\{\omega_1\}$) such that for some φ, $P = \|X_i\varphi\|^{M^e}$. $\{\omega_1\}$ is in $\mathcal{A}_i(\omega)$ whenever $\{\omega_2, \omega_3\}$ is, and Ω always belongs to $\mathcal{A}_i(\omega)$. Finally, *ari* is equivalent in this case to *arx*, since $\mathcal{X}_i(\omega)$ is already closed under supersets of $\bigcap \mathcal{X}_i(\omega)$. □

Furthermore, at every world in the model, Iggy is *rational* in at least the sense that his beliefs satisfy:

K$_X$. $(X_i(\varphi \to \psi) \wedge X_i\varphi) \to X_i\psi$
D. $X_i\varphi \to \neg X_i\neg\varphi$
N. $X_i\top$

K_X is stronger than we need to call Iggy rational, of course, but its satisfaction straightforwardly implies the more generally plausible rationality axioms:

K$_{AX}$. $(X_i(\varphi \to \psi) \wedge X_i\varphi) \to (A_i\psi \to X_i\psi)$
CON$_X$. $(X_i\varphi \wedge X_i\psi) \to X_i(\varphi \wedge \psi)$

Together these mean that there's no difference between Iggy's implicit and explicit beliefs—Iggy has drawn every classical logical consequence that can be drawn from what he believes given everything of which he is aware.

Thus, let $\boldsymbol{\mathcal{M}}^{\dagger\dagger}$ designate the restriction of $\boldsymbol{\mathcal{M}}^{\dagger}$ models to those where the $\mathcal{X}_i$ satisfy the following constraints (for all i, ω, P_1, P_2):

kax. If $P_1 \in \mathcal{X}_i(\omega)$ and $P_1 \subseteq P_2$, then $P_2 \in \mathcal{X}_i(\omega)$ if $P_2 \in \mathcal{A}_i(\omega)$
conx. $\bigcap \mathcal{X}_i(\omega) \in \mathcal{X}_i(\omega)$
d. If $P \in \mathcal{X}_i(\omega)$, then $\Omega \setminus P \notin \mathcal{X}_i(\omega)$
n. $\Omega \in \mathcal{X}_i(\omega)$

Then the model also obviously establishes:

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

That is, even under the strengthened awareness conditions (*arx*, *ara* and *ari*) which generalise PLA^b and AUR, *plus* basic rationality conditions (*kax*, *conx*, *d* and *n*), there are going to PWA models 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.

Consequently, where the goal is just to develop a logic of non-veridical belief with unawareness, the possible worlds theorist has at least one sure way out of Dekel *et al.*'s triviality result—and they can rest easy that there will be no further triviality results in the nearby vicinity, either.

10 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 approach to content. This sets much the work on unawareness somewhat at odds with research elsewhere in formal epistemology, where coarse-grained models of propositional content are still very much standard—and for good reason.

Moreover, the present state of affairs leaves the possible worlds theorist 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 coarse-grained content model of *belief* with unawareness, and indeed we can include stronger axioms that generalise the intuitions behind PLA and AUR without introducing triviality. 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. Whether we ought to keep both of those axioms, on the other hand, is a trickier matter.

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.

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 $\Gamma = \{\varphi_1, \ldots, \varphi_n\}$ is consistent iff $\varphi_1 \wedge \cdots \wedge \varphi_n$ is consistent; and an arbitrary set $\Gamma = \{\varphi_1, \varphi_2, \ldots\}$ 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 [16]. 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 define the canonical PWA model, M^c:

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

1. $\Omega^c = \Omega^0 \cup \Omega^1$
2. $\mathcal{X}_i^c(\Gamma^n) = \begin{cases} \emptyset, & \text{if } \Gamma^n \backslash A_i = \emptyset \\ \{|\varphi| : X_i\varphi \in \Gamma^n\} \cup \{\Omega^1, \Omega^0\}, & \text{if } \Gamma^n \backslash A_i \neq P^\star_{\alpha,\Gamma^n} = \emptyset \\ \{|\varphi| : X_i\varphi \in \Gamma^n\} \cup \{P^\star_{\alpha,\Gamma^n}\}, & \text{if } P^\star_{\alpha,\Gamma^n} \neq \emptyset \end{cases}$
3. $\mathcal{A}_i^c(\Gamma^n) = \begin{cases} \emptyset, & \text{if } \Gamma^n \backslash A_i = \emptyset \\ \mathfrak{B}(\{|\varphi| : A_i\varphi \in \Gamma^n\} \cup 2^{\Omega^1}), & \text{if } \Gamma^n \backslash A_i \neq \emptyset \end{cases}$
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 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, $|\neg\varphi| = \Omega^c \setminus |\varphi|$. That $\{|\varphi| : A_i\varphi \in \Gamma^n\}$ is closed under binary intersections follows a basically similar structure, using CON_A and Lemma 1. □

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* $\Gamma^n \backslash A_i \neq \emptyset$

Proof The right-to-left direction is trivial. To establish the left-to-right direction, we suppose throughout that $\Gamma^n \backslash A_i \neq \emptyset$, and hence $\{|\varphi| : A\varphi \in \Gamma^n\} \neq \emptyset$. 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 $P \in \{|\varphi| : A_i\varphi \in \Gamma^n\}$
c_2 $P \subseteq \Omega^1$
c_3 $P \cap \Omega^0 = P' \cap \Omega^0$, 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 $P \cap \Omega^0 = \emptyset$, 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 $\{\omega_1\}$ satisfies c_2, then $\{\omega_1\} \cap P_2$ satisfies c_2 (for any P_2)
(d) If $\{\omega_1\}$ and P_2 both satisfy c_1, then $\{\omega_1\} \cap P_2$ satisfies c_1
(e) If $\{\omega_1\}$ and c_2 both satisfy c_3, then $\{\omega_1\} \cap P_2$ satisfies c_3
(f) If $\{\omega_1\}$ satisfies c_1, and P_2 satisfies c_3, then $\{\omega_1\} \cap P_2$ 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 $P = \emptyset$, 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 $\mathcal{X}_i^c$ satisfy *pla*: For the case when $\mathcal{A}_i^c = \emptyset$, $\mathcal{X}_i^c = \emptyset$ by definition. For the cases where $\mathcal{A}_i^c$ is non-empty, note that when $P^\star_{\alpha,\Gamma^n} = \emptyset$, then Ω^1 and Ω^0 are in both $\mathcal{X}_i^c(\Gamma^n)$ and $\mathcal{A}_i^c(\Gamma^n)$; and when $P^\star_{\alpha,\Gamma^n} \neq \emptyset$, $P^\star_{\alpha,\Gamma^n} \in \mathcal{X}_i^c(\Gamma^n)$ 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 $\mathcal{X}_i^c$ 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 $\|\psi\|^{M^c} \in \mathcal{X}_i^c(\Gamma^n)$ iff $|\psi| \in \mathcal{X}_i^c(\Gamma^n)$. Suppose that $M^c, \Gamma^n \models X_i\psi$. By Definition 2, it follows that $\|\psi\|^{M^c} = |\psi| \in \mathcal{X}_i^c(\Gamma^n)$. 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 $P^\star_{\alpha,\Gamma^n} = \emptyset$ (since $\emptyset \subseteq P^\star_{\alpha,\Gamma^n} \subseteq \Omega^1$). So by Definition 3, if $|\psi| \in \mathcal{X}_i^c(\Gamma^n)$, 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 $|\psi| = \|\psi\|^{M^c} \in \mathcal{X}_i^c(\Gamma^n)$, 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,

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

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, $A_i\neg(\varphi \wedge \neg\gamma) \in \Gamma^n$. Since we can also show that awareness is closed under logical equivalence, so $A_i(\varphi \to \gamma) \in \Gamma^n$. Putting these points together, in Γ^n we have:

$$\begin{array}{l} A_i\psi, \\ A_i(\varphi_n \to \psi), \\ A_i\big(\varphi_{n-1} \to (\varphi_n \to \psi)\big) \\ \quad\vdots \\ A_i\big(\varphi_1 \to \big(\ldots(\varphi_n \to \psi)\ldots\big)\big) \end{array}$$

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 [8, pp. 59ff] for details.

Appendix B

For Theorem 3, I'll just sketch a proof that $\Sigma \cup \{\mathrm{AR}_X\}$ 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 $\Sigma \cup \{\mathrm{AR}_X\}$. The proof is then the same as for Theorem 1, with the following addition to Lemma 4:

That M^c satisfies *arx*: Let $Q = \{\Delta^i : P \in \mathcal{X}_j^c(\Delta^i)\}$. We need to show that if $Q \in \mathcal{A}_i^c(\Gamma^n)$, then $P \in \mathcal{A}_i^c(\Gamma^n)$. Assume that $Q \in \mathcal{A}_i^c(\Gamma^n)$. 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, $P \in \mathcal{A}_i^c(\Gamma^n)$, 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 $\mathcal{X}_j^c(\Delta^1) = \mathcal{X}_j^c(\Delta^0)$, so $\Delta^1 \in Q$ iff $\Delta^0 \in Q$. By points already established (Lemma 3), then, $Q \in \mathcal{A}_i^c(\Gamma^n)$ 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 $P \notin (2^{\Omega^1} \setminus \{\emptyset\}) \cup \{\Omega^0\}$, P will be in $\mathcal{X}_j^c(\Delta^i)$ 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 AR_X, $A_i\psi \in \Gamma^n$; so $|\psi| \in \{|\varphi| : A_i\varphi \in \Gamma^n\}$, which is of course a subset of $\mathcal{A}_i^c(\Gamma^n)$.

Acknowledgements Thanks are due to [*anonymised*].

References

1. Braddon-Mitchell, D., Jackson, F.: Philosophy of Mind and Cognition. Blackwell, Malden (1996)
2. Bradley, R.: Decision Theory with a Human Face. Cambridge University Press, Cambridge (2017)
3. Bueno, O., Menzel, C., Zalta, E.N.: Worlds and Propositions Set Free. Erkenntnis **79**, 797–820 (2014)
4. Camp, E.: Thinking with Maps. Philosophical Perspectives **21**, 145–182 (2007)
5. Chalmers, D.: Does Conceivability Entail Possibility? In: T. Gendler, J. Hawthorne (eds.) Conceivability and Possibility, pp. 145–200. Oxford University Press, Oxford (2002)
6. Chalmers, D.: The Foundations of Two-Dimensional Semantics. In: M. Garcia-Carpintero, J. Macia (eds.) Two-Dimensional Semantics, pp. 55–140. Oxford University Press (2006)
7. Chalmers, D.: The Nature of Epistemic Space. In: A. Egan, B. Weatherson (eds.) Epistemic Modality, pp. 60–107. Oxford University Press, Oxford (2011)
8. Chellas, B.F.: Modal Logic: An Introduction. Cambridge University Press, Cambridge (1980)
9. Chen, Y.C., Ely, J.C., Luo, X.: Note on unawareness: Negative Introspection versus AU Introspection (and KU Introspection). International Journal of Game Theory **41**, 325–329 (2012)
10. Cozic, M.: Probabilistic Unawareness. Games **7**(4), 38 (2016)
11. Cummins, R.: Inexplicit Information. In: M. Brand, R.M. Harnish (eds.) The Representation of Knowledge. University of Arizona Press, Tucson (1986)
12. Cummins, R.: Why Adding Machines Are Better Examples than Thermostats: Comments on Dretske's 'The explanatory role of content'. In: Contents of Thought: Proceedings of the 1985 Oberlin Colloquium in Philosophy. University of Arizona Press, Tucson (1987)
13. Davies, M.: Meaning, Quantification, Necessity: Themes in Philosophical Logic. Routledge & Kegan Paul, London (1981)
14. Dekel, E., Lipman, B.L., Rustichini, A.: Standard State-Space Models Preclude Unawareness. Econometrica **66**, 159–173 (1998)
15. Egan, A.: Seeing and Believing: perception, belief formation and the divided mind. Philosophical Studies **140**(1), 47–63 (2008)

16. Fagin, R., Halpern, J.Y.: Belief, Awareness, and Limited Reasoning. Artificial Intelligence **34**, 39–76 (1988)
17. Field, H.H.: Mental Representation. Erkenntnis **13**, 9–61 (1978)
18. Fodor, J.: The Language of Thought. Cromwell, New York (1975)
19. Fodor, J.: Psychosemantics: The problem of meaning in the philosophy of mind. MIT Press (1987)
20. Fritz, P., Lederman, H.: Standard State Space Models of Unawareness. In: Proceedings TARK 2015. DOI 10.4204/EPTCS.215.11. URL https://arxiv.org/abs/1606.07520
21. Halpern, J.Y.: Alternative Semantics for Unawareness. Games and Economic Behavior **37**, 321–339 (2001)
22. Halpern, J.Y., Rêgo, L.C.: Reasoning about knowledge of unawareness. Games and Economic Behavior **67**(2), 503–525 (2009)
23. Heifetz, A., Meier, M., Schipper, B.C.: Interactive Unawareness. Journal of Economic Theory **130**, 78–94 (2006)
24. Heifetz, A., Meier, M., Schipper, B.C.: A Canonical Model for Interactive Unawareness. Games and Economic Behavior **62**, 304–324 (2008)
25. Hintikka, J.: Knowledge and Belief: An introduction to the logic of the two notions. Cornell University Press, Ithaca (1962)
26. Hofweber, T.: Are there ineffable aspects of reality? In: K. Bennett, D. Zimmerman (eds.) Oxford Studies in Metaphysics, Vol. 10. Oxford Unviersity Press, Oxford (2016)
27. Jackson, F.: From Metaphysics to Ethics. Oxford University Press (1998)
28. Jackson, F.: Possibilities for representation and credence: two space-ism versus one space-ism. In: A. Egan, B. Weatherson (eds.) Epistemic Modality. Oxford University Press, Oxford (2009)
29. Kaplan, D.: A Problem in Possible Worlds Semantics. In: W. Sinnott-Armstrong, D. Raffman, N. Asher (eds.) Modality, Morality and Belief: Essays in Honor of Ruth Barcan Marcus, pp. 41–52. Cambridge University Press, Cambridge (1995)
30. Lewis, D.: Radical interpretation. Synthese **27**(3), 331–344 (1974)
31. Lewis, D.: Logic for Equivocators. Nous **16**(3), 431–441 (1982)
32. Lewis, D.: New work for a theory of universals. Australasian Journal of Philosophy **61**(4), 343–377 (1983)
33. Lewis, D.: On the Plurality of Worlds. Cambridge University Press (1986)
34. Lewis, D.: Reduction of Mind. In: S. Guttenplan (ed.) Companion to the Philosophy of Mind, pp. 412–431. Blackwell (1994)
35. Li, J.: Information Structures with Unawareness. Journal of Economic Theory **144**, 977–993 (2009)
36. Modica, S., Rustichini, A.: Awareness and partitional information structures. Theory and Decision **37**(1), 107–124 (1994)
37. Modica, S., Rustichini, A.: Unawareness and Partitional Information Structures. Games and Economic Behavior **27**, 265–298 (1999)
38. Montague, R.: Universal Grammar. Theoria **36**, 373–398 (1970)
39. Nagel, T.: The view from nowhere. Oxford University Press (1986)
40. Pitt, D.: Mental representation. In: E.N. Zalta (ed.) The Stanford Encyclopedia of Philosophy, winter 2018 edn. Metaphysics Research Lab, Stanford University (2018)
41. Sandgren, A.: Secondary belief content, what is it good for? Philosophical Studies **175**(6), 1467–1476 (2018)
42. Schipper, B.C.: Awareness-dependent Subjective Expected Utility. International Journal of Game Theory **42**, 725–753 (2013)
43. Schipper, B.C.: Preference-based unawareness. Mathematical Social Sciences **70**, 34–41 (2014)
44. Schipper, B.C.: Awareness. In: H. van Ditmarsch, J.Y. Halpern, W. van der Hoek, B. Kooi (eds.) Handbook of Epistemic Logic, pp. 77–146. College Publications, London (2015)
45. Schwarz, W.: Against Magnetism. Australasian Journal of Philosophy **92**(1), 17–36 (2014)
46. Schwitzgebel, E.: A Phenomenal, Dispositional Account of Belief. Nous **36**(2), 249–275 (2002)
47. Scott, D.: Advice in Modal Logic. In: K. Lambert (ed.) Philosophical Problems in Logic. Reidel (1970)
48. Sillari, G.: Models of Awareness. In: G. Bonanno, W. van der Hoek, M. Woolridge (eds.) Logic and the Foundations of Game and Decision Theory: Proceedings of the Seventh Conference (2008)

49. Sillari, G.: Quantified Logic of Awareness and Impossible Possible Worlds. The Review of Symbolic Logic **1**(4), 514–529 (2008)
50. Stalnaker, R.: Assertion. In: P. Cole (ed.) Pragmatics, vol. 9, pp. 78–95. New York Academic Press, New York (1978)
51. Stalnaker, R.C.: Inquiry. The MIT Press, London (1984)
52. Stalnaker, R.C.: The Problem of Logical Omniscience, I. Synthese **89**(3), 425–440 (1991)
53. Uzquiano, G.: Modality and Paradox. Philosophy Compass **10**(4), 284–300 (2015)
54. Velázquez-Quesada, F.R.: Explicit and Implicit Knowledge in Neighbourhood Models. In: G. D., R. O., H. H. (eds.) International Workshop on Logic, Rationality and Interaction, pp. 239–252. Springer
55. Walker, O.: Unawareness with "possible" possible worlds. Mathematical Social Sciences **70**, 23–33 (2014)
56. Williamson, T.: Knowledge and its Limits. Oxford University Press, Oxford (2002)
57. Wittgenstein, L.: Tractatus Logico-Philosophicus. Kegan Paul, Trench, Trubner (1922)
58. Yalcin, S.: Belief as Question-Sensitive. Philosophy and Phenomenological Research **97**(1), 23–47 (2018)