Journal of Mechanism and Institution Design
ISSN: 2399-844X (Print), 2399-8458 (Online)
DOI:10.22574/jmid.2022.12.001

OBJECT-BASED UNAWARENESS: THEORY AND APPLICATIONS

Oliver J. Board

Paul | Weiss, USA
ojboard@paulweiss.com

Kim-Sau Chung

Hong Kong Baptist University, China
kschung@hkbu.edu.hk

ABSTRACT

In this paper and its companion paper, Board & Chung (2021), we provide foundations for a model of unawareness that can be used to distinguish between what an agent is unaware of and what she simply does not know. At an informal level, this distinction plays a key role in a number of recent papers such as Tirole (2009) and Chung & Fortnow (2016). Here we provide a set-theoretic (i.e., non-linguistic) version of our framework. We use our *object-based unawareness structures* to investigate two applications. The first application provides a justification for the *contra proferentem doctrine* of contract interpretation, under which ambiguous terms in a contract are construed against the drafter. Our second application examines speculative trade.

Keywords: Unawareness, legal doctrine, no-trade theorem.

JEL Classification Numbers: D83, D86, D91, K12.

This paper was first circulated in 2008. The literature has since grown much bigger than is reflected in our references. We apologize for not being able to do justice to this subsequent literature. We thank Eddie Dekel, Lance Fortnow, Joseph Halpern, Jing Li, Ming Li, and seminar participants at various universities for very helpful comments. We also thank Zaifu Yang for inviting us to submit the paper to this *Journal*. All errors are ours.

1. INTRODUCTION

THERE are two strands of literature on unawareness and it seems that they are unaware of each other.

The first unawareness literature (let's call it the *applied* literature) consists of applied models, such as Tirole (2009) and Chung & Fortnow (2016), where agents are uncertain whether they are aware of everything that their opponents are aware of, and have to strategically interact under these uncertainties. For example, in Tirole (2009), a buyer and a seller negotiate a contract as in the standard hold-up problem. At the time of negotiation, there may or may not exist a better design for the product. Even if a better design exists, however, the contracting parties may not be aware of it. If a party is aware of it, he can choose whether or not to point it out to the other party. But even if he is not aware of it, he *is* aware that a better design may exist and his opponent may be aware of this better design. In Tirole's words, "parties are unaware, but aware that they are unaware"; and they have to negotiate under this uncertainty. Chung & Fortnow (2016) consider the plight of an American founding father drafting a Bill of Rights that will be interpreted by a judge 200 years later. The founding father is aware of some rights, but is uncertain whether or not there are other rights that he is unaware of. Here, as in Tirole (2009), the founding father is unaware, but aware that he may be unaware; and he has to decide how to write the Bill of Rights under this uncertainty.

The second unawareness literature (let's call it the *foundational* literature) attempts to provide a more rigorous account of the properties of unawareness: see, e.g., Fagin & Halpern (1987), Modica & Rustichini (1994), Modica & Rustichini (1999), Dekel et al. (1998), Halpern (2001), Li (2009), Halpern & Rêgo (2006), Sillari (2006), and Heifetz et al. (2006), Heifetz et al. (2013). These authors are motivated by the concern that *ad hoc* applied models, if not set up carefully enough, may go awry in the sense that agents in those models may violate rationality in some way, as captured by various introspection axioms first articulated in Modica & Rustichini (1994) and Dekel et al. (1998) (which we shall refer to as the DLR axioms hereafter).[1] The rest of this literature proposes various models that are set up carefully enough to take these concerns into account.

[1] In particular, two of the key DLR axioms are *KU-introspection* ("the agent cannot know that he is unaware of a specific event") and *AU-introspection* ("if an agent is unaware of an event E, then he must be unaware of being unaware of E").

These two literatures are somewhat disconnected. For example, Tirole makes no reference to any work in the foundational literature, nor does he explain whether or not his agents satisfy the DLR axioms that are the main concerns of that literature. Similarly, none of the studies in the foundational literature explains whether Tirole's model fits in their framework, and if not, whether Tirole's agents violate some or all of the DLR axioms. This paper and Board & Chung (2021) attempt to connect these two literatures.

There is a reason why it is difficult to directly compare Tirole's model with the majority of the models proposed in the foundational literature. To propose a model and to provide foundations for it, an author needs to explain how her model should be interpreted. This is typically done by showing how her model assigns truth conditions to each sentence in a particular formal language; i.e., by the procedure of systematically giving yes/no answers to a laundry list of questions such as: "at state w, does agent i know that it is sunny in New York?"[2] Note, however, the formal language chosen by the author defines the laundry list of questions she is ready to give yes/no answers to. A question not expressible in her chosen formal language is hence not a legitimate question. The answer to it is neither yes nor no—she simply is not ready to say.

Unfortunately, questions such as "at state w, is agent i aware that he is not aware of everything?" are *not* expressible in the formal languages chosen by many authors in the foundational literature (notable exceptions include Halpern & Rêgo (2006) and Sillari (2006), which we shall return to shortly). The formal languages chosen by these authors do not contain quantifiers such as "everything", thus rendering "aware of everything" an inexpressible concept. In other words, while in Tirole's model, "parties are unaware, but aware that they are unaware", it is difficult to tell whether this is also true of the agents in most of the models proposed in the foundational literature. The answer is neither yes nor no—these authors simply are not ready to say.

Several contributions to the foundational literature, mostly coming from logicians and computer scientists, do work with formal languages that contain quantifiers; see, e.g., Halpern & Rêgo (2006) and Sillari (2006). Their proposed models, however, look very different from applied economic models used in, for example, Tirole (2009) and Chung & Fortnow (2016). For

[2] In many of the studies more familiar to economists (see e.g., Li (2009)), although this procedure is not performed explicitly, there is still a clear way to assign truth conditions within an appropriately-specified formal language according to the author's description of her proposed model.

example, in the model proposed by Halpern & Rêgo (2006), there is a *syntactic* awareness function that assigns to every state and every agent a set of sentences in their chosen formal language. The interpretation is that this set is the set of facts that the agent is aware of at that state. This "list of sentences" approach to construct models is very flexible, but may be deemed unhelpful by economists. This may explain why this approach, while not uncommon among logicians, is rarely seen in economics.[3]

In the specific case of Halpern & Rêgo (2006), there is a deeper reason why their proposed model is *not* the same as the models used in the applied literature. Recall that in the latter models, although agents know what they are aware of, they may be uncertain whether or not they are aware of everything. Such uncertainty cannot arise in the model proposed by Halpern & Rêgo (2006), however.[4]

To summarize, while the assumption that "agents are unaware, but are aware that they are unaware" plays a key role in much of the applied literature of unawareness, the foundations of these models remain unclear. We do not know whether agents in these models violate some or all of he DLR axioms that are the main concerns of the foundational literature. This paper and Board & Chung (2021) aim to provide this missing foundation.

In these two papers, we describe a model, or more precisely a class of

[3] To provide an analogy that may help elucidate this comparison, consider the difference between Aumann's information partition model, where a partition of the state space is used to encode an agent's knowledge of events, and a "list of sentences" approach where knowledge is instead modeled by a list of sentences describing exactly what that agent knows.

[4] For readers who are familiar with Halpern & Rêgo (2006), this can be proved formally as follows. Recall the following definition in Halpern & Rêgo (2006): "*Agents know what they are aware of* if, for all agents i and all states s,t such that $(s,t) \in \mathscr{K}_i$ we have that $\mathscr{A}_i(s) = \mathscr{A}_i(t)$." So it suffices to prove that, in any instance of Halpern & Rêgo (2006) structure, if there is a state t such that agent i is uncertain whether or not there is something he is unaware of, then there must be another state s such that $(s,t) \in \mathscr{K}_i$ but $\mathscr{A}_i(s) \neq \mathscr{A}_i(t)$. Let $\alpha = \exists x \neg A_i x$ represent "there is something that agent i is unaware of". Therefore, $\neg\alpha$ means "there is nothing that agent i is unaware of". Let $\beta = A_i\alpha \wedge A_i\neg\alpha \wedge \neg X_i\alpha \wedge \neg X_i\neg\alpha$ represent "agent i is aware of both α and $\neg\alpha$ but he does not know whether α or $\neg\alpha$ is true (recall that X_i is Halpern & Rêgo (2006)'s explicit knowledge operator). In short, β means "agent i is uncertain whether or not there is something he is unaware of". Let M be any instance of Halpern & Rêgo (2006)'s structure, and t is a state such that $(M,t) \models \beta$. Then we have $(M,t) \models \neg K_i\alpha \wedge \neg K_i\neg\alpha$ (recall that K_i is Halpern & Rêgo (2006)'s implicit knowledge operator). Therefore, there exists a state s such that $(t,s) \in \mathscr{K}_i$ and $(M,s) \models \neg\alpha$, and another state s' such that $(t,s') \in \mathscr{K}_i$, and $(M,s') \models \alpha$. Since $\alpha = \exists x \neg A_i x$, there exists ϕ such that $\phi \in \mathscr{A}_i(s)$ and $\phi \notin \mathscr{A}_i(s')$. But that means at least one $\mathscr{A}_i(s)$ and $\mathscr{A}_i(s')$ is different from $\mathscr{A}_i(t)$.

models, called *object-based unawareness structures* (OBU structures). Readers will find that these structures encompass models used in the applied literature. In comparison with the applied literature, however, we provide complete and rigorous foundations for these structures. The formal language we choose to work with is rich, and in particular contains quantifiers, enabling us to describe explicitly whether or not agents are aware that they are unaware. We provide an axiomatization for these structures and verify that all of the DLR axioms are satisfied. The value of thinking about agents who exhibit this kind of uncertainty has already been demonstrated by the existing applied literature; we demonstrate the tractability of our framework by considering further applications.

A key feature of our structures is that unawareness is object-based: A seller may be unaware of a better design, or a founding father may be unaware of a particular right. In contrast, in models of unforeseen contingencies, agents cannot foresee every contingency, or every state. This raises the question of whether the agents in our structures are aware of every state. We do not have an answer to this question. As we explained above, our understanding of any proposed model is constrained by the formal language we choose to work with. Although we have already chosen to work with a formal language much richer than most in the foundational literature, there are still questions that fall outside of it. We do not have answers to these questions, simply because we do not speak that language.

The division of labor between this paper and Board & Chung (2021) is as follows. In Board & Chung (2021), we give the model-theoretic description of OBU structures by showing how they assign truth conditions to every sentence of a formal language. We then prove a model-theoretic soundness and completeness theorem, which characterizes OBU structures in terms of a system of axioms. We then verify that agents in OBU structures do not violate any of the DLR axioms that are generally considered to be necessary conditions for a plausible notion of unawareness. Board & Chung (2021) also contain a more complete literature review, as well as a discussion of several variants of OBU structures.

In this paper, we give a set-theoretic description of the OBU structures. Although less formal than the model-theoretic treatment, we hope this will be more accessible to the general audience. In parallel to the model-theoretic soundness and completeness theorem in Board & Chung (2021), we prove set-theoretic completeness results in this paper.

The second half of this paper considers two applications. First, we use the model to provide a justification for the *contra proferentem doctrine* of contract interpretation, commonly used to adjudicate ambiguities in insurance contracts. Under *contra proferentem*, ambigous terms in a contract are construed against the drafter. Our main result is that when the drafter (the insurer) has greater awareness than the other party (the insured), *and when the insured is aware of this asymmetry*, *contra proferentem* minimizes the chances that the insured forgoes gain of trade for fear of being exploited. On the other hand, when there is no asymmetric awareness, efficiency considerations suggest no reason to prefer *contra proferentem* over an alternative interpretive doctrine that resolves ambiguity in favor of the drafter.

From the perspective of our theory, an argument common among legal scholars as far back as Francis Bacon, that *contra proferentem* encourages the insurer to write clearer contracts, misses the point. If a more precise contract increases the surplus to be shared between the insurer and the insured, market forces provide incentives to draft such a contract regardless of the interpretive doctrine employed by the court. The advantage of *contra proferentem* is rather that it enables the insurer to draft more acceptable contracts, by expanding the set of events that he can credibly insure.

Our second application examines speculative trade. We first generalize the classical No Trade Theorem to situations where agents are delusional but nevertheless act so as to satisfy a weaker condition called terminal partitionality. We then introduce the concepts of *living in denial* (i.e., agents believe, perhaps incorrectly, that there is nothing that they are unaware of) and *living in paranoia* (i.e., agents believe, perhaps incorrectly, that there is something that they are unaware of). We show that both living in denial and living in paranoid, in the absence of other forms of delusion, imply terminal partitionality, and hence the no trade theorem result obtains.

The structure of this paper is as follows. Section 2 describes our OBU structures, and Section 3 shows how to incorporate probabilities. Section 4 presents the first application, and Section 5 the second. Section 6 concludes.

2. OBU STRUCTURES

In thi section we introduce OBU structures and present set-theoretic completeness results[5] that provide a precise characterization of the properties of knowledge, unawareness etc. For the sake of transparency, and to aid interpretation, we also include in Appendix A the model-theoretic description of these structures; i.e., we show how OBU structures assign truth conditions for a formal language (a version of first-order modal logic).

2.1. Modeling knowledge and unawareness

An *OBU structure* for n agents is a tuple $\langle W, O, \{O_w\}, \{\mathscr{I}_i\}, \{\mathscr{A}_i\}\rangle$, where:

- W is a set of states;
- O is a set of objects;
- $O_w \subseteq O$ is the set of objects that really exist at state w;
- $\mathscr{I}_i : W \to 2^W$ is an *information function* for agent i; and
- $\mathscr{A}_i : W \to 2^O$ is an *awareness function* for agent i.

Intuitively, $\mathscr{I}_i(w)$ indicates the states that agent i considers possible when the true state is w, while $\mathscr{A}_i(w)$ indicates the objects she is aware of. The sets O_w will not be used until we describe quantified events in section 2.3 below.

In the standard information partition model familiar to economists, events are represented as subsets of the state space, corresponding to the set of states in which some given proposition is true. In OBU structures, we try to carry around one more piece of information when we represent an event, namely the set of objects referred to in the verbal description of that event. Formally, an event is an ordered pair (R, S), where $R \subseteq 2^W$ is a set of states and $S \subseteq 2^O$ is a set of objects; we call R the *reference* of the event (denoted by $ref(R, S)$), corresponding (as before) to the set of states in which the proposition is true; and S is the *sense* of the event (denoted by $sen(R, S)$), listing the set of objects referred to in the proposition. (To give an example, the events representing the propositions "the dog barked" and "the dog barked and the cat either did

[5] This purely semantic approach to epistemic logic was pioneered by Halpern (1999).

or did not meow" have the same reference but difference senses.) We sometimes abuse notation and write (R,a) instead of $(R,\{a\})$, and (w,S) instead of $(\{w\},S)$. We use $\mathscr{E}$ to denote the set of all events, with generic element E.

We now define two operators on events, corresponding to "not" and "and":

$$\begin{aligned} \neg(R,S) &= (W \setminus R, S), \\ \wedge_j \left(R_j, S_j\right) &= \left(\cap_j R_j, \cup_j S_j\right). \end{aligned}$$

The negation of an event holds at precisely those states at which the event does not hold, but it refers to the same set of objects. The conjunction of several events holds only at those states at which *all* of those events hold, and it refers to each set of objects. It will often be convenient to use disjunction ("or") as well, defined in terms of negation and conjunction as follows:

$$\begin{aligned} \vee_j \left(R_j, S_j\right) &= \neg\left(\wedge_j \neg\left(R_j, S_j\right)\right) \\ &= \left(\cup_j R_j, \cup_j S_j\right). \end{aligned}$$

In OBU structures, there are three modal operators for each agent, representing awareness, implicit knowledge, and explicit knowledge:

$$\begin{aligned} \mathsf{A}_i(R,S) &= (\{w \mid S \subseteq \mathscr{A}_i(w)\}, S) \text{ (awareness)} && (1)\\ \mathsf{L}_i(R,S) &= (\{w \mid \mathscr{I}_i(w) \subseteq R\}, S) \text{ (implicit knowledge)} && (2)\\ \mathsf{K}_i(R,S) &= \mathsf{A}_i(R,S) \wedge \mathsf{L}_i(R,S) \text{ (explicit knowlege)} && (3) \end{aligned}$$

Intuitively, an agent is aware of an event at w if she is aware of every object in the *sense* of the event; and the agent implicitly knows an event at state w if the *reference* of the event includes every state she considers possible. However, implicit knowledge is not the same as explicit knowledge, and the latter is our ultimate concern. Implicit knowledge is merely a benchmark that serves as an intermediate step to modeling what an agent actually knows. Intuitively, an agent does not actually (i.e., explicitly) know an event unless he is aware of the event *and* he implicitly knows the event. Notice that A_i, L_i, and K_i do not change the set of objects being referred to.

It is easy to verify that awareness and implicit knowledge satisfy the following properties (where we suppress the agent-subscripts):

A1 $\wedge_j \mathsf{A}\left(R, S_j\right) = \mathsf{A}\left(R, \cup_j S_j\right)$

A2 $\mathsf{A}(R,S) = \mathsf{A}(R',S)$ for all R, R'

A3 $\mathsf{A}(R,\varnothing) = (W,\varnothing)$

A4 $\mathsf{A}(R,X) = (R',X)$ for some R'

L1 $\mathsf{L}(W,O) = (W,O)$

L2 $\wedge_j \mathsf{L}\left(R_j,S\right) = \mathsf{L}\left(\cap_j R_j,S\right)$

L3 $\mathsf{L}(R,S) = (R',S)$ for some R'

L4 if $\mathsf{L}(R,S) = (R',S)$ then $\mathsf{L}(R,S') = (R',S')$

The following results show that L1–L4 and A1–A4 also provide a precise characterization of awareness and implicit knowledge, respectively.

Proposition 1. *Suppose that A_i is defined as in (1). Then:*

1. *A_i satisfies A1–A4; and*
2. *if A'_i is an operator on events which satisfies A1–A4, we can find an awareness function $\mathscr{A}_i$ such that A'_i and A_i coincide.*

Proposition 2. *Suppose that L_i is defined as in (2). Then:*

1. *L_i satisfies L1–L4; and*
2. *if L'_i is an operator on events which satisfies L1–L4, we can find an information function $\mathscr{I}_i$ such that L'_i and L_i coincide.*

The proofs of these and all other results can be found in the appendix.

2.2. Introducing Properties

In an OBU structure, we take as primitives not individual events such as “John is tall”, but rather individual *properties* such as “... is tall”. Intuitively, the property “... is tall” can be thought of as a correspondence from objects to states, telling us for each object at which states it possesses this property. More generally, properties can be represented as functions from objects to events: $p : O \to \mathscr{E}$ such that

$$p(a) = \left(R^p_a, S^p \cup \{a\}\right) \text{ for some } R^p_a \subseteq W \text{ and some } S^p \subseteq O.$$

Intuitively, R_a^p is the set of states where object a possesses property p, and S^p is the set of objects referred to in the description of the property; for example, if p is the property "... is taller than Jim", then $S^p = \{Jim\}$. Note that S^p could be the empty set, for example if p is the property "... is tall". Let $\mathscr{P}$ denote the class of all these functions.

REMARK: In many applications, such as the one we will study in Section 4, the set of properties that are relevant to the problem at hand is a much smaller set than $\mathscr{P}$, and hence not every (R,S) pair is a representation of a proposition like "John is tall".

REMARK: Although we have only described 1-place properties, this is without loss of generality, because we can build up n-place properties from n 1-place properties. Suppose we want to construct the 2-place property $taller(a,b)$, to be interpreted as "a is taller than b". We start with a family of 1-place properties $\{p_a : O \to \mathscr{E}\}_{a \in O}$, to be interpreted "$a$ is taller than ...". Define $f : O \to \mathscr{P}$ as $f(a) = p_a$. Then the two-place property $taller : O^2 \to \mathscr{E}$ is defined by taller $(a,b) = f(a)(b)$. Notice that, in particular, the sense of the event $taller(a,b)$ is $\{a,b\}$, because

$$sen(f(a)(b)) = S^{f(a)} \cup \{b\} = \{a\} \cup \{b\}.$$

We can also take negations, conjunctions, and disjunctions of properties:

$$\begin{aligned}
\neg p &: O \to \mathscr{E} \text{ such that } (\neg p)(a) = \neg(p(a)) \\
p \wedge q &: O \to \mathscr{E} \text{ such that } (p \wedge q)(a) = p(a) \wedge q(a) \\
p \vee q &: O \to \mathscr{E} \text{ such that } (p \vee q)(a) = p(a) \vee q(a)
\end{aligned}$$

We also use $p \to q$ as shorthand for $\neg p \vee q$.

REMARK: It is worth noting that the concept of negation defined above does not coincide with the everyday English notion of "opposites" (as in "short is the opposite of tall"). There are two reasons for this: first, even if we restrict attention to people (humans), we might argue some people are neither tall nor short (for instance, an white male who is 5 foot tall); second, there are objects which are neither tall nor short simple because they don't have a height at all (for instance, an abstract object such as "a thought". Therefore we prefer to think of tall and short as two separate properties, allowing for the possibility that short is not the same as not tall.

2.3. Quantified Events

In many applications, we want to deal not only with events such as "a is a better design" and "agent i knows that a is a better design", but also events such as "agent i is not aware of any better design" and "agent i does not know whether there is a better design that he is unaware of". These events involve *quantification*. In this section, we show how they are handled in OBU structures.

To begin with, we should note that everyday English admits multiple interpretations of quantifiers (such as the word "all"), corresponding to different scopes implicit in the conversation: the "universe of objects" referred to by the word "all" can vary. We often freely switch back and forth among different interpretations, without making the scope explicit, and leaving it for the context to resolve the ambiguity. In a formal model, however, these different interpretations must be explicitly distinguished by different quantifiers. Two particular quantifiers that may get confused are the *possibilitist quantifier* and the *actualist quantifier*; the former has a scope that spans all possible objects, while the latter has a scope that spans only those objects that really exist at a given state. The quantifier that is used in OBU structures is the actualist one.

To illustrate the difference between these two quantifiers, consider the following application. Suppose we want to model Hillary's uncertainty regarding whether or not Bill has an illegitimate child. The simplest way to do it is to have Hillary consider as possible two different states, w_1 and w_2, but Bill's illegitimate child really exists at only one of these states. Using a to denote "Bill's illegitimate child", it means $a \in O_{w_1} \subset O$ but $a \notin O_{w_2}$. Since Hillary cannot tell apart these two states, she does not know for sure whether Bill has an illegitimate child or not. However, such a simple model of Hillary's uncertainty "works" only because the existential quantifier used by this simple model is the actualist one. If a reader misinterprets the model as using the possibilitist quantifier, he would have regarded it as a poor model of Hillary's uncertainty: "Since Bill's illegitimate child 'exists' at every state that Hillary considers possible, Hillary knows for sure that Bill has an illegitimate child, and hence there is no uncertainty at all!"

We define possibilitist-quantified events first, because they are simpler, and can be used as an intermediate step to define actualist-quantified events. For any property $p \in \mathscr{P}$, let $\overline{\mathsf{All}}\, p$ denote the event that "all objects satisfy property p", where "all" is interpreted in the possibilitist sense. Formally, $\overline{\mathsf{All}}$

is a mapping from properties to events, such that

$$\overline{\mathsf{All}}\, p = \left(\cap_{a\in O} R_a^p, S^p\right).$$

So $\overline{\mathsf{All}}\, p$ holds at precisely those worlds where $p(a)$ is true for each objects a in the universal set O, and it refers only to those objects referred to by property p.

We defined actualist-quantified events, or simply *quantified events*. First recall that an OBU structure specifies, for each state w, the set $O_w \subseteq O$ of objects that really exist at that state. We define a special property *re* ("... is real") in terms of these sets:

$$re(a) = \left(\{w \mid a \in O_w\}, a\right). \tag{4}$$

Let $\mathsf{All}\, p$ denote the event that "all objects satisfy property p", where "all" is interpreted in the actualist sense. Formally, All is a mapping from properties to events, such that

$$\mathsf{All}\, p = \left(\cap_{a\in O} R_a^{re\to p}, S^p\right). \tag{5}$$

Intuitively, $\mathsf{All}\, p$ holds at every state where all real objects possess property p; and the sense of $\mathsf{All}\, p$ is precisely the objects used to describe property p. It is easy to verify that the actualist quantifier satisfies the following properties:

All**1** $\mathsf{All}\left(\wedge_j p_j\right) = \wedge_j \left(\mathsf{All}\, p_j\right)$

All**2** if $w \in R_a^p$ for every $a \in O$, then $w \in ref\left(\mathsf{All}\, p\right)$

All**3** if $R_a^p = R_a^q$ for every $a \in O$, then $ref\left(\mathsf{All}\, p\right) = ref\left(\mathsf{All}\, p\right)$

All**4** $sen\left(\mathsf{All}\, p\right) = S^p$

The following result shows that All1 – All4 also provide a precise characterization of the actualist quantifier.

Proposition 3. *Suppose that* All *is defined as in (4) and (5). Then:*

1. All *satisfies* All1 – All4*; and*
2. *if* All′ *is a mapping from properties to events which satisfies* All1 – All4*, we can find a collection of real objects* $\{O_w\}$ *such that* All′ *and* All *coincide.*

3. OBU STRUCTURES WITH PROBABILITIES

It is easy to introduce probabilistic beliefs into the OBU structures, although Board & Chung (2021)'s axiomatization does not include this part. We first introduce implicit beliefs, once again as a benchmark case that serves as an intermediate tool to modeling what the agent actually believes. The relation between explicit beliefs (i.e., an agent's actual beliefs) and implicit beliefs is then analogous to the relation between explicit knowledge and implicit knowledge.

Let us begin with an OBU structure $\langle W, O, \{O_w\}, \{\mathscr{I}_i\}, \{\mathscr{A}_i\}\rangle$. To avoid unnecessary complications, let's assume that W is finite. Augment the OBU structure with $\{q_i\}_{i\in N}$, where each q_i is a probability assignment that associates with each state w a probability distribution on W satisfying $q_i(w)(\mathscr{I}_i(w)) = 1$ (i.e., an agent (implicitly) assigns probability 1 to those states that he considers possible when the true state is w). For any real number r, we introduce two belief operators for each agent, mapping any given event $E = (R, S) \in \mathscr{E}$ to the events that an agent implicitly and explicitly, respectively, believes that E holds with probability at least r:

$$\overline{\mathsf{B}}_i^r(R,S) = (\{w \mid q_i(w)(R) \geq r\}, S) \text{ (implicit belief)} \tag{6}$$

$$\mathsf{B}_i^r(R,S) = \mathsf{A}_i(R,S) \wedge \overline{\mathsf{B}}_i^r(R,S) \text{ (explicit belief).} \tag{7}$$

An *augmented OBU structure* is a tuple $\langle W, O, \{O_w\}, \{\mathscr{I}_i\}, \{\mathscr{A}_i\}, \{q_i\}\rangle$.

The common prior assumption is considered controversial, even in the absence of unawareness (Morris, 1995; Gul, 1998). Nevertheless, to facilitate comparison with the existing literature in Section 5, we introduce it here. We say that an augmented OBU structure satisfies the *common prior assumption* if there exists a probability distribution q on W such that, whenever $q(\mathscr{I}_i(w)) > 0$, we have

$$q_i(w)(\cdot) = q(\cdot \mid \mathscr{I}_i(w)),$$

where $q(\cdot \mid \mathscr{I}_i(w))$ is the conditional probability distribution on W given $\mathscr{I}_i(w)$. When an augmented OBU structure satisfies the common prior assumption, we can represent it as the tuple $\langle W, O, \{O_w\}, \{\mathscr{I}_i\}, \{\mathscr{A}_i\}, q\rangle$, and simply call it an OBU structure with common prior.

4. THE CONTRA PROFERENTEM DOCTRINE

Verba fortius accipiuntur contra proferentem (literally, "words are to be taken most strongly against him who uses them") is a rule of contractual interpretation which states that ambiguities[6] in a contract should be construed against the party who drafted the contract. This rule (henceforth *cp doctrine*) finds clear expression in the *First Restatement of Contracts*[7] (1932) as follows:

> Where words or other manifestations of intention bear more than one reasonable meaning an interpretation is preferred which operates more strongly against the party from whom they proceed, unless their use by him is prescribed by law.

Although the principles for resolving ambiguity are more nuanced in the *Second Restatement* (1979), the cp doctrine is widely applied in the context of insurance contracts; indeed, Abraham (1996) describes it as "the first principle of insurance law".

In this section, we use OBU structures to formalize the rationale behind this rule. In particular, we compare it with the opposite doctrine that resolves ambiguity *in favor of* the drafter. We first show that there is a form of symmetry between these two doctrines, and neither systematically outperforms the other if there is no asymmetric unawareness. We then introduce asymmetric unawareness and explain in what sense the cp doctrine is a superior interpretive doctrine.

Let an OBU structure with common prior $\langle W, O, \{O_w\}, \{\mathscr{I}_i\}, \{\mathscr{A}_i\}, q\rangle$ be given.[8] Assume that there are two agents. Agent 1 is a (female) risk-neutral

[6] "Ambiguity" is an ambiguous term in economics, and often refers to situations where decision makers entertain multiple prior probability distributions. Here, we are referring to the layman's use of the word, that is to a situation where language is susceptible to multiple interpretations.

[7] The *Restatements of the Law* are treatises published by the American Law Institute as scholarly refinements of black-letter law, to "address uncertainty in the law through a restatement of basic legal subjects that would tell judges and lawyers what the law was." Although non-binding, the authoritativeness of the Restatements is evidenced by their near-universal acceptance by courts throughout the United States.

[8] Given our earlier comments about the common prior assumption, the reader may wonder why we impose this assumption here. The common prior assumption allows us to state our results neatly. But we otherwise do not believe that the comparison between different doctrines depends on this assumption.

insurer and agent 2 is a (male) risk-averse insured. In the absence of any insurance contract between the agents, agent 1's income is \$0 in every world, while agent 2's income is \$0 in some worlds and \$1 in other worlds. We can think of 0 income as the result of some negative income shock, which the risk-averse agent 2 would like to insure against. Agent 1's utility is equal to her income, and agent 2's utility is $U(\cdot)$, which is strictly increasing and strictly concave in his income.

One of the elements in O, denoted by ι, is agent 2's income. (We will explain what else is contained in O later.) Let $Z \subsetneq W$ be the (nonempty) set of states in which agent 2 suffers an income shock. The event "agent 2 suffers an income shock" is hence $E = (Z, \iota)$. It is natural to assume that agent 2 is always aware of his own income (i.e., $\iota \in \mathscr{A}_2(w)$ for every w), and so agent 2 can always form an explicit probabilistic belief about event E (given by $q(ref(E))$=$q(Z)$).

To make the setup as noncontroversial as possible, we make a couple of standard assumptions:

1. Each agent i's $\mathscr{I}_i$ forms a partition of the state space W; i.e., $w \in \mathscr{I}_i(w)$ for every $w \in W$, and $w' \in \mathscr{I}_i(w)$ implies $\mathscr{I}_i(w') = \mathscr{I}_i(w)$.

2. Each agent i (implicitly) knows what he is aware of; i.e., $w' \in \mathscr{I}_i(w)$ implies $\mathscr{A}_i(w') = \mathscr{A}_i(w)$.

We also make an additional assumption motivated by the current application:

3. Agent 1 is aware of more objects than agent 2 is: $\mathscr{A}_2(w) \subseteq \mathscr{A}_1(w)$ for every $w \in W$.

The third assumption captures the idea that agent 1 (the insurer) is the more sophisticated party in this transaction. In what follows we analyze a special case that satisfies these assumptions: agent 1 is aware of everything while agent 2 is aware of nothing except his own income: $\mathscr{A}_1(w) = O$ and $\mathscr{A}_2(w) = \{\iota\}$ for all w; and both agents are completely uninformed: $\mathscr{I}_i(w) = W$ for all w and $i = 1, 2$. This allows us to abstract away from the classical adverse selection problem, which is already well understood, and focus instead on the interaction between contractual ambiguity and asymmetric awareness.

Note that, although we make the extreme assumption that agent 2 is aware of nothing (except his own income), we do not preclude that he is aware of his own unawareness. For example, as long as $O_w \setminus \{\iota\} \neq \varnothing$ for all w, the event "agent 1 is aware of something that agent 2 is unaware of" (where "some" is interpreted in the actualist sense) is the event $(W,\varnothing)$. Since

$$\mathsf{K}_2(W,\varnothing) = \mathsf{A}_2(W,\varnothing) \wedge \mathsf{L}_2(W,\varnothing) = (W,\varnothing) \wedge (W,\varnothing) = (W,\varnothing), \quad (8)$$

agent 2 *explicitly* knows that "agent 1 is aware of something that agent 2 is unaware of" in every state w.

If we further assume that $O_w = \hat{O} \subset O$ for all w, then agent 2 knows how many objects there are that agent 1 is aware of but agent 2 is not. Although this assumption is not realistic (even if the insured is certain that there are *some* objects that he is unaware of, he will typically be uncertain about the exact number of such objects), it simplifies the analysis considerably. In this preliminary investigation of the cp doctrine, therefore, we add this assumption. To further simplify, we assume that $\hat{O} = O$ until section 4.3.2 where it becomes important to distinguish the two sets.

The timing of the contracting game is as follows. In stage one, agent 1 proposes an insurance contract. The contract specifies a premium, a payment, and the circumstances under which agent 1 (the insurer) has to pay the insurance payment to agent 2 (the insured). A critical assumption is that the payout circumstances have to be described in an exogenously given language, to be defined shortly, and cannot make reference to agent 2's income. Without this assumption, the insurance problem would be trivial. This assumption makes sense when, for example, agent 2's income is not verifiable and hence not contractible, or if contracting on income would create a serious moral hazard problem. In stage two, agent 2 either accepts the contract and pays the premium, or rejects it. If he accepts, we move to stage three, the contract enforcement stage, where nature randomly picks a state according to the probability distribution q, and agent 1 has to pay agent 2 the insurance payment unless she can prove to a court that the payout circumstances do not obtain.

4.1. Contracts and Interpretations

We now define the *contractual language*, which is built up from the following elements (the *vocabulary*):

- $a,b,c\dots$ — an exogenously given, nonempty list of (names of) objects, which together with agent 2's income ι form the the set O in our OBU structure (i.e., $O \setminus \{\iota\} = \{a,b,c,\dots\}$).

- $P_1, P_2, \ldots$ — an exogenously given, nonempty list of *predicates*, each of which will later on be construed (by the court) as corresponding to a specific property.[9]

- $\neg$ (not), $\wedge$ (and), $\vee$ (or) — Boolean operators.

Note that by identifying the set of objects' names with the objects themselves, we are assuming that there is no ambiguity in the interpretation of these names; we make the simplifying assumption that all contractual ambiguity relates to which properties the various predicates stand for.

Formally, the contractual language is a collection of sentences, each of which is a finite string of *letters* (i.e., elements of the vocabulary) satisfying a certain grammatical structure. We define this collection recursively as follows:

(i) for each object a and predicate P, $P(a)$ (to be interpreted as "object a is P") is a sentence;

(ii) if ϕ and ψ are sentences, then $\neg\phi$, $\phi \wedge \psi$, and $\phi \vee \psi$ are sentences.

The contractual language, denoted by $\mathscr{L}$, is the smallest set satisfying (i) and (ii).[10] If b and r are objects and F and L are predicates, an example of a sentence in $\mathscr{L}$ is $F(b) \wedge L(r)$, with a possible interpretation of "the basement is flooded and the roof is leaking".

An *insurance contract* is a triple (g, h, ϕ), where $g \in \mathbb{R}_+$ is the insurance premium that agent 2 pays agent 1 *ex ante*, and $\phi \in \mathscr{L}$ is a sentence that describes the circumstances under which agent 1 pays $h \in \mathbb{R}_+$ to agent 2 *ex post*.

Although a predicate P (in the vocabulary of the contractual language) is supposed to correspond to a specific property, whether an object satisfies that property or not is often ambiguous *ex post*. For example, consider a health insurance contract that covers the cost of a hip replacement just when it is medically necessary. Is a patient who is able to walk, but only with a great deal of pain, covered? Some people might say yes, while others would say no.

---

[9] Without loss of generality, we assume that all these predicates are 1-place. See Section 2 for discussion.

[10] We could further expand our contractual language to include quantifiers. We conjecture that this would not affect our main results.

Without this kind of ambiguity, the cp doctrine would be moot. So we now introduce this kind of ambiguity into our model.

We capture this kind of ambiguity by supposing that there may be disagreement about which property (in an OBU structure) a given predicate corresponds to. Formally, an *interpretation* is a mapping l from predicates to properties. To keep things simple, imagine that there are two sub-populations of society, and each has its own interpretation of every predicate P. Let l_1 and l_2 denote these two interpretations. It is natural to assume that $S^{l_1(P)} = S^{l_2(P)}$.

An interpretation l that maps predicates to properties can be extended to a mapping from the contractual language $\mathscr{L}$ to events in the obvious way:

I1 $l(P(a)) = \left(R_a^{l(P)}, S^{l(P)} \cup \{a\}\right)$;

I2 $l(\neg\phi) = \neg l(\phi)$;

I3 $l(\phi \wedge \psi) = l(\phi) \wedge l(\psi)$;

I4 $l(\phi \vee \psi) = l(\phi) \vee l(\psi)$.

We can now formalize the cp doctrine. The cp doctrine instructs the court to resolve any ambiguity against the party who drafted the contract (i.e., agent 1 in this model). In the example above, if the hip replacement is medically necessary given one interpretation but not the other, then under cp doctrine the court should rule in favor of agent 2 and require agent 1 to payout. Formally, the cp doctrine is a mapping from $\mathscr{L}$ to events given by

$$d_{cp}(\phi) = l_1(\phi) \vee l_2(\phi) \quad \text{for all } \phi \in \mathscr{L}.$$

Note that d_{cp} is *not* an interpretation, since it may not satisfy I2 or I3.

For sake of comparison, we set up a strawman and define the mirror image of the cp doctrine, the *anti-cp doctrine*, which instructs the court to resolve any ambiguity in favor of agent 1. Formally, $d_{anti-cp}$ is given by

$$d_{anti-cp}(\phi) = l_1(\phi) \wedge l_2(\phi) \quad \text{for all } \phi \in \mathscr{L}.$$

The interpretive doctrine of the court is commonly known. Given this interpretive doctrine d, agent 1's problem in stage three (the contract enforcement stage) is to prove to the court that the payout circumstances do not obtain, or equivalently that event $d(\phi)$ has not happened.

We assume that, once the true state w is realized, agent 1 has sufficient evidence to prove that object a satisfies property p *if and only if* (1) a is real ($a \in O_w$), and (2) a does in fact satisfy property p ($w \in R_a^P$). Under our earlier simplifying assumption that $O_w = \hat{O} = O$ for every w, condition (1) is always satisfied.

Finally, we need to explain how agent 2 evaluates a given contract and makes his accept/reject decision accordingly in stage two. This can be tricky, as it depends on how agent 2's awareness changes after he reads the contract (which may mention objects that agent 2 was unaware of before he read it). We postpone this discussion to section 4.3 below, and first consider a benchmark case where there is symmetric awareness between the two agents. The central message from the benchmark case is this: linguistic ambiguity alone (without asymmetric unawareness) is not sufficient to justify the cp doctrine.

EXAMPLE: Let's use an example to illustrate our setup. Consider the simplest case where there is only one object name, a, and one predicate, P, in the contractual language. One can think of a as "the basement", and P as "... is flooded". Suppose there are only two states: w_1 and w_2. At w_1, there is a lot of water in the basement, and everyone in the society would agree that the basement is flooded. But at w_2, the basement is merely wet, and not everyone in the society would think that it is flooded. Therefore we have $l_1(P(a)) = (\{w_1, w_2\}, a)$ and $l_2(P(a)) = (\{w_1\}, a)$. Suppose the contract says that the insured will be compensated when the basement is flooded; i.e., the contract takes the form of $(g, h, P(a))$. Under the cp-doctrine, the insured will be compensated at both states; whereas under the anti-cp doctrine, he will be compensated only at state w_1. As another example, suppose the contract says that the insured will be compensated when the basement is *not* flooded; i.e., the contract takes the form of $(g, h, \neg P(a))$. Under the cp-doctrine, the insured will be compensated at state w_2; whereas under the anti-cp doctrine, he will never be compensated.

4.2. Benchmark: Symmetric Awareness

Before we continue the description of our model, let's first consider the benchmark case of *symmetric awareness*, where $O_1(w) = O_2(w) = O$ for every $w \in W$. In this case, agent 2 is aware of every object that agent 1 is aware of. Since both agents are aware of every object, implicit knowledge/beliefs and explicit knowledge/beliefs coincide. This reduces our model back to a standard exercise in contract theory. The introduction of an exogenous contractual language does not pose a new methodological challenge, because its only effect is to restrict the contracting parties' ability to approximate a first-best contract. Different interpretive doctrines imply different restrictions on the contracting parties. However, as we shall see shortly, there is a strong symmetry between the restrictions implied by the cp doctrine and those implied by the anti-cp doctrine, and hence no systematic advantage for the former over the latter.

A first best contract is any contract that requires the insurer to pay \$1 to the insured exactly in those states where he suffers an income shock.[11] Recall that Z denotes the set of states where the insured suffers an income shock. Since the contracting parties cannot write contracts that directly refer to agent 2's income, they have to look for (contractible) events that correlate with agent 2's income shock. In other words, they have to look for a $\phi \in \mathscr{L}$ such that, under a given interpretive doctrine d, the set $ref(d(\phi))$ approximates Z. How well $ref(d(\phi))$ approximates Z depends on the prior probability q; or, more precisely, on $q(ref(d(\phi)) \setminus Z)$ and $q(Z \setminus ref(d(\phi)))$.

To make this more precise, let $\mathscr{R}_{cp} = \{ref(d_{cp}(\phi)) \mid \phi \in \mathscr{L}\}$ denote the set of references that can be described under the cp doctrine; similarly, let $\mathscr{R}_{anti-cp} = \{ref(d_{anti-cp}(\phi)) \mid \phi \in \mathscr{L}\}$. Then say that the cp doctrine *systematically out-performs* the anti-cp doctrine if and only if $\mathscr{R}_{anti-cp} \subsetneq \mathscr{R}_{cp}$.

To see that this definition captures the correct intuition, suppose first that $\mathscr{R}_{anti-cp} \nsubseteq \mathscr{R}_{cp}$. Then there is some (non-empty)[12] $R \in \mathscr{R}_{anti-cp} \setminus \mathscr{R}_{cp}$. If $Z = R$ and q is the uniform prior, then full insurance is possible only under the anti-cp doctrine. On the other hand, if $\mathscr{R}_{anti-cp} \subsetneq \mathscr{R}_{cp}$, any insurance outcome achievable under the anti-cp doctrine can be replicated under the cp doctrine, while we can find a case where full insurance is possible only under the cp doctrine.

EXAMPLE CONTINUED: Let's use our earlier example to illustrate what is at stake when the society chooses between the two doctrines. In that example,

$$\mathscr{R}_{cp} = \{\varnothing, \{w_2\}, \{w_1, w_2\}\}.$$

Note that the singleton set $\{w_1\}$ is not in $\mathscr{R}_{cp}$. Therefore, full insurance is not always possible under the cp doctrine. In particular, if $Z = \{w_1\}$ (i.e., the in-

[11] The insurance premium is a pure transfer and hence has no efficiency implications.

[12] It is easy to see that $\varnothing \in \mathscr{R}_{anti-cp} \cap \mathscr{R}_{cp}$.

sured's wealth drop is correlated with how severely his basement is flooded), the contractual language would be found inadequate for the purpose of providing insurance—in fact, the optimal insurance contract will be no insurance in such an unfortunate case. Now, consider the counterfactual case where the parties anticipate that the court would interpret their contract using the anti-cp doctrine. Under such anticipation, they can sign a contract of the form $(g,h,P(a))$; and with $d_{anti-cp}(P(a)) = (\{w_1\},a) = (Z,a)$, perfect insurance can be achieved. But does it mean that the anti-cp doctrine is better than the cp doctrine? The answer is no, because by a symmetric argument we can see that, in case $Z = \{w_2\}$, perfect insurance can be achieved under the cp doctrine but not under the anti-cp doctrine. Without further information regarding which case is more likely, it is impossible to rank the two doctrines.

The following proposition says that $|\mathscr{R}_{anti-cp}| = |\mathscr{R}_{cp}|$, and so it cannot be the case that the cp doctrine systematically outperforms the anti-cp doctrine.

Proposition 4. $|\mathscr{R}_{anti-cp}| = |\mathscr{R}_{cp}|$.

Proof. It suffices to show that $R \in \mathscr{R}_{anti-cp}$ if and only if $W \setminus R \in \mathscr{R}_{cp}$. Suppose $R \in \mathscr{R}_{anti-cp}$. Then there exists $\phi \in \mathscr{L}$ such that $ref(d_{anti-cp}(\phi)) = R$. But $\phi \in \mathscr{L}$ implies $\neg\phi \in \mathscr{L}$. Since $ref(d_{cp}(\neg\phi)) = ref(l_1(\neg\phi) \vee l_2(\neg\phi)) = ref(\neg l_1(\phi) \vee \neg l_2(\phi)) = ref(\neg l_1(\phi)) \cup ref(\neg l_2(\phi)) = (W \setminus ref(l_1(\phi))) \cup (W \setminus ref(l_2(\phi))) = W \setminus (ref(l_1(\phi)) \cap ref(l_2(\phi))) = W \setminus ref(l_1(\phi) \wedge l_2(\phi)) = W \setminus ref(d_{anti-cp}(\phi)) = W \setminus R$, we have $W \setminus R \in \mathscr{R}_{cp}$. The other direction is similar. $\square$

We find it illuminating to contrast Proposition 4 with an argument common among legal scholars as far back as Francis Bacon, that the advantage of *contra proferentem* is to provide incentives for the insurer to write precise contracts. Carolina Care Plan Incorporated v. McKenzie (2007) provides a succinct statement of this argument in a recent ruling: "Construing ambiguity against the drafter encourages administrator-insurers to write clear plans that can be predictably applied to individual claims, countering the temptation to boost profits by drafting ambiguous policies and construing them against claimants." However, in light of Propositon 4, this argument misses the point. According to our theory, more precise contracts will be rewarded by higher premiums regardless of the interpretative doctrine employed by the court.

4.3. Asymmetric Awareness

We now return to the case of asymmetric awareness: $\mathscr{A}_1(w) = O$ and $\mathscr{A}_2(w) = \{\iota\}$ for all $w \in W$. Here, an important modelling question to address is: how would agent 2's awareness changes after he reads a contract which mentions objects that he was previously unaware of?

If agent 2 was unaware of those objects because they slipped his mind, then it would be natural to assume that he becomes aware of them once he reads about them in the contract. If, instead, he was unaware of them because he genuinely had no idea what they were, then it would be more natural to assume that his awareness would not change even after reading the contract. In reality there would likely be some objects in each category, which begs a richer model that distinguishes a slip-the-mind object from a genuinely-clueless object. For the sake of simplicity, we keep the two cases distinct and analyze each in turn.

Although the the slip-the-mind case is not the only case where unawareness can arise, it is the only case that has been considered by other authors so far.[13] However, in the current setup, it turns out that the slip-the-mind case and the benchmark case (with symmetric awareness) generate the same outcome. Hence linguistic ambiguity, even when coupled with unawareness, is not sufficient justification for the cp doctrine, if the unawareness is of the slip-the-mind variety. In the genuinely-clueless case, on the other hand, we show that a case can be made in favor of the CP doctrine.

4.3.1. The Slip-the-Mind Case

When agent 2 reads a contract that mentions objects that he was previosuly unaware of, and if he was unaware of them simply because they slipped his mind, he will become aware of those objects after he reads the contract. Suppose the contract is (g, h, ϕ). Let S be the set of objects mentioned in the sentence ϕ; i.e., $S = sen(l_1(\phi)) = sen(l_2(\phi)) = sen(d(\phi))$ for both interpretive doctrines d. Before agent 2 reads the contract, his awareness function is $\mathscr{A}_2(w) = \{\iota\}$ for all w; after he reads the contract, his awareness function becomes $\mathscr{A}_2(w) = \{\iota\} \cup S$ for all w.

Recall that $E = (Z, \iota)$ is the event "agent 2 suffers an income shock". So

[13] See, for example, Filiz-Ozbay (2012), Ozbay (2007), and Tirole (2009).

the four events

$$E \wedge d(\phi), \quad E \wedge \neg d(\phi), \quad \neg E \wedge d(\phi), \quad \neg E \wedge \neg d(\phi),$$

that are relevant for agent 2's accept/reject decision all have the same *sense*, namely $\{\iota\} \cup S$. Since after reading the contract, $\mathscr{A}_2(w) = \{\iota\} \cup S$ for every w, agent 2 can form explicit probabilistic beliefs about these events. This allows him to calculate the expected utilities resulting from accepting and rejecting the contract.

A simple backward induction argument then suggests that the insurer, who is aware of every object throughout, will choose a $\phi \in \mathscr{L}$ such that $ref(d(\phi))$ best approximates Z, and internalizes the gains from trade by setting the insurance premium at the level that makes agent 2 indifferent between accepting and rejecting. As in the benchmark case, the insurer's ability to approximate an arbitrary Z is restricted by the contractual language, and the exact restrictions depend on the interpretive doctrine d. This is captured by the fact that both $\mathscr{R}_{cp}$ and $\mathscr{R}_{anti-cp}$ are in general strictly smaller than 2^W.

By Proposition 4, we know that $|\mathscr{R}_{anti-cp}| = |\mathscr{R}_{cp}|$, so neither doctrine systematically outperforms the other. Either $\mathscr{R}_{anti-cp} = \mathscr{R}_{cp}$ (in which case the choice of the interpretive doctrine is irrelevant), or $\mathscr{R}_{anti-cp} \setminus \mathscr{R}_{cp} \neq \varnothing$ (in which case one can readily construct an example where full insurance is possible only under the anti-cp doctrine).

4.3.2. *The Genuinely-Clueless Case*

To help understand the clueless case, consider the example of a pet insurance policy. Such policies typically list the various diseases that are covered by the policy.[14] The list contains diseases such as balanoposthitis, esophagitis, enteritis, enucleation, FIP, HGE, hemobartonella, histiocytoma, leptospirosis, neoplasia, nephrectomy, pneumothorax, pyothorax, rickettsial, tracheobronchitis .... Most insureds have no idea what these diseases are even *after* reading the insurance contract. This is exactly what we assume in the clueless case, where agent 2's awareness function is the same before as after reading the contract; i.e., $\mathscr{A}_2(w) = \{\iota\}$ for all w.

A knee-jerk intuition may suggest that no contract with a positive premium will be accepted by agent 2, because he cannot fully understand it. "If

[14] See, for example, the policies offered at www.petinsurance.com.

I am offered a contract that reads ($10, $100, "Barney catches disease xxx")," the knee-jerk intuition argues, "then the chances are that Barney will never catch xxx, and the insurer will never need to pay me anything." We shall see shortly that the knee-jerk intuition is half right but also half wrong. Understanding why it is half wrong is the key to understanding why the cp doctrine is the superior interpretive doctrine.

Consider two different insurance policies, one covering balanoposthitis but not tracheobronchitis, and the other covering tracheobronchitis but not balanoposthitis. These two policies clearly differ, but the insured would not be able to base his accept/reject decision on the basis of this difference if he unaware of both diseases. Suppose he knows that some diseases are common and expensive to treat, while others are rare and inexpensive to treat. If the insured takes into account that the insurance policy is written by a rational insurer, who in turn knows that the insured is unaware of either disease, then a simple game-theoretic argument would enable the insured to figure out that the disease covered in the actual contract he is offered must be the less expensive one. Note that agent 2's pessimism does not follow logically from unawareness per se, but rather from the analysis of his opponent's strategic behavior.

This informal argument suggests that we can analyze the clueless case by representing it as an imperfect information game. Agent 1's actions are the different contracts she can write. Agent 2 does not perfectly observe agent 1's action. But those actions are partitioned into different information sets for agent 2. A contract that covers only balanoposthitis belongs to the same information set as another contract that covers only tracheobronchitis (assuming it has the same premium and payment as the first one), and both are in a different information set from a third contract that covers both balanoposthitis and tracheobronchitis, which in turn belongs to the same information set as a fourth contract that covers leptospirosis and brucellosis, and so on. In any (perfect Bayesian) equilibrium of such a game, agent 2 must hold pessimistic beliefs with respect to any information set on the equilibrium path.

Let's illustrate this idea using a simple example, which also serves to counter the knee-jerk intuition above.

In this simple example, l_1 is the same as l_2, so there is no linguistic ambiguity and the choice of interpretive doctrine is irrelevant (we are merely trying to demonstrate that some insurance is possible even under asymmetric unawareness). So there is no need to distinguish predicates and properties. There are three states: $W = \{w_1, w_2, w_3\}$. Agent 2 suffers an income shock in states w_1 and w_2: $E = (\{w_1, w_2\}, \iota)$. There are infinitely many objects: $O = \{\iota, a, b, c, \ldots\}$, but $O_w = \hat{O} = \{\iota, a, b\}$ for all w. There is only one predicate/property: P, with $P(a) = (\{w_1, w_2\}, a)$, $P(b) = (w_1, b)$, and $P(x) = (\varnothing, x)$ for $x = c, d, \ldots$ As stated above, we assume that $\mathscr{I}_1(w) = \mathscr{I}_2(w) = W$, $\mathscr{A}_1(w) = O$, and $\mathscr{A}_2(w) = \{\iota\}$ for all w. The prior q puts equal probability on each state.

In this example, agent 2 explicitly knows that agent 1 is aware of some objects that he is unaware of; indeed, he explicitly knows that the number of such objects is exactly two (see the discussion following equation (8) above). He explicitly knows that there exists *something* that satisfies property P most of the time, although he is unaware of what it is. He also explicitly knows that there exists something else that satisfies property P less often, but at least whenever that *something* satisfies P he will also suffer an income shock. More importantly, he explicitly knows that there does not exist anything that never satisfies P. Thus when he sees a contract of the form $(g, 1, P(\cdot))$, where g satisfies

$$3U(1-g) \geq 2U(1) + U(0), \tag{9}$$

he will be willing to accept the contract even though he is unaware of the specific object mentioned in the contract. In equilibrium, the insurer will offer the contract $(g^*, 1, P(b))$ such that g^* satisfies (9) with equality.[15]

The above example is a counter-example to the knee-jerk intuition. Although it is natural to think of the set O as being very large,[16] $\hat{O}$ need not be, or at least agent 2 need not believe that it is. If agent 2 believes that there are not that many things that he is unaware of, he would be less worried about being tricked. The initial appeal of the knee-jerk intuition comes from an implicit assumption that $\hat{O}$ is big. We shall call this the rich-object assumption, and formalize it as follows. For any sentence $\phi \in \mathscr{L}$, the events $l_1(\phi)$, $l_2(\phi)$, $d_{cp}(\phi)$, and $d_{anti-cp}(\phi)$ all have the same (nonempty) sense, call it S. Suppose $S = \{a_1, \ldots, a_n\}$, and write ϕ as $\phi[a_1, \ldots, a_n]$ to make this explicit. From any

[15] It is important to understand why the insurer will not offer, for instance, the contract $(g^*, 1, P(c))$, even though such a contract will also be accepted by the insured. There is no real object that bears the name "c" that the insurer can point to to prove to the court that $P(c)$ does not obtain; given that the burden of proof in on the insurer to show that he does not have to payout, he will have to payout in every state.

[16] O is the set of hypothetical as well as real objects, and hence is limited only by our agents' imagination.

sentence $\phi[a_1,\ldots,a_n]$, and any n distinct objects $b_1,\ldots,b_n$, we can construct another sentence $\phi[b_1,\ldots,b_n]$ which is the same as $\phi[a_1,\ldots,a_n]$ with each a_j replaced by b_j. It is easy to verify that $\phi[b_1,\ldots,b_n]$ is also an element of $\mathscr{L}$.

Assumption 5 (The Rich-Object Assumption). *Let d denote the interpretive doctrine used by the court. For any sentence $\phi[a_1,\ldots,a_n] \in \mathscr{L}$, either $ref(d(\phi[a_1,\ldots,a_n])) = W$, or there exist n distinct objects, $b_1,\ldots,b_n$, such that*

1. *$b_1,\ldots,b_n \in \hat{O}$, and*

2. *$ref(d(\phi[b_1,\ldots,b_n])) = \varnothing$.*

Note that the Rich-Object Assumption is a joint assumption on $\hat{O}$ and the interpretive doctrine d: fixing $\mathscr{L}$, l_1, and l_2, $\hat{O}$ may satisfy the Rich-Object Assumption under one doctrine d but not under another. The importance of the Rich-Object Assumption is summarized by the following proposition, the first part of which formalizes the knee-jerk intuition.

Proposition 6. *Let d denote the interpretive doctrine used by the court.*

1. *If the Rich-Object Assumption holds, then in any perfect Bayesian equilibrium, agent 2 receives no insurance.*

2. *If the Rich-Object Assumption does not hold, then there exists nonempty $R \subseteq W$ such that, if agent 2 suffers an income shock exactly in states in R, then there exists a perfect Bayesian equilibrium where agent 1 offers a contract that fully insures agent 2, and agent 2 accepts it.*

Proof. 1. Suppose $(g,h,\phi[a_1,\ldots,a_n])$ is a contract that is both offered and accepted with positive probability in any equilibrium.
If $ref(d(\phi[a_1,\ldots,a_n])) = W$, then the fact that it is offered with positive probability in equilibrium implies that $h \le g$, and hence agent 2 receives no insurance under this contract. Suppose $ref(d(\phi[a_1,\ldots,a_n])) \subsetneq W$. Then $(g,h,\phi[b_1,\ldots,b_n])$, where $\phi[b_1,\ldots,b_n]$ is as defined in the Rich-Object Assumption, will also be accepted with positive probability. However, by the Rich-Object Assumption, agent 1 can always prove that the event $d(\phi[b_1,\ldots,b_n])$ does not obtain and hence avoid paying the insurance premium h. The fact that the original contract is offered with

positive probability implies that agent 1 also never needs to pay the insurance premium under that contract. Hence agent 2 receives no insurance from it.

2. Let $\phi[a_1,\ldots,a_n]$ be a sentence that invalidates the Rich-Object Assumption. Let $(b_1^*,\ldots,b_n^*)$ be a solution of the following minimization problem:

$$\min_{\underbrace{b_1,\ldots,b_n}_{distinct}\in\hat{O}} q(ref(d(\phi[b_1,\ldots,b_n]))),$$

where the existence of a solution is guaranteed by the finiteness of W. Finally, define R to be $ref(d(\phi[b_1^*,\ldots,b_n^*]))$. By assumption, R is nonempty. Then, if agent 2 suffers an income shock exactly in states in R, contracts of the form $(g,1,\phi[b_1^*,\ldots,b_n^*])$ will fully insure agent 2. A simple argument then establishes the existence of a perfect Bayesian equilibrium where agent 1 offers this contract with the insurance premium g such that agent 2 is indifferent between accepting and rejecting, and agent 2 accepts the contract. The fact that $(b_1^*,\ldots,b_n^*)$ solves the above minimization problem implies that agent 1 cannot profitably deviate to other contracts within the equivalence class of $\{(g,1,\phi[b_1,\ldots,b_n]) \mid b_1,\ldots,b_n \text{ distinct}\}$. □

We can now formalize the benefit of the cp doctrine over the anti-cp doctrine: the cp doctrine minimizes the chance that the Rich-Object Assumption holds.

Proposition 7. *Whenever the Rich-Object Assumption holds under the cp doctrine, it will also hold under the anti-cp doctrine.*

Proof. It suffices to observe that, for any $\phi\in\mathscr{L}$, $ref(d_{anti-cp}(\phi))\subseteq ref(d_{cp}(\phi))$. □

The converse of Proposition 7 is not true, as illustrated by the following simple example.

EXAMPLE: In this example, there are two states, $W=\{w_1,w_2\}$, two contractible objects, a and b, and one predicate, P. The two interpretations of P are as follows:

$$\begin{aligned} l_1(P(a)) &= (w_1,a), & l_1(P(b)) &= (w_2,b),\\ l_2(P(a)) &= (\varnothing,a), & l_2(P(b)) &= (W,b). \end{aligned}$$

Suppose $Z = \{w_1\}$. Then, under the cp-doctrine, agent 1 can offer a contract $(g, h, P(a))$, with appropriately g and h, and fully insures agent 2. (Full insurance is achieved because $d_{cp}(P(a)) = (w_1, a)$.) Even when agent 1 anticipates that agent 2 will accept both contracts $(g, h, P(a))$ and $(g, h, P(b))$, as he cannot distinguish the two, she will have no incentive to deviate to offering contract $(g, h, P(b))$, because $d_{cp}(P(a)) = (W, a)$. The same is not true under the anti-cp doctrine. Indeed, it is a mechanical exercise to check that the Rich-Object Assumption is satisfied under the anti-cp doctrine. For example, if agent 1 anticipates that agent 2 will accept the contract $(g, h, P(b))$, she will deviate to contract $(g, h, P(a))$, because $d_{anti-cp}(P(b)) = (w_2, b)$, while $d_{anti-cp}(P(a)) = (\varnothing, b)$. Similarly, if agent 1 anticipates that agent 2 will accept the contract $(g, h, P(b) \wedge \neg P(a))$, she will deviate to contract $(g, h, P(a) \wedge \neg P(b))$, because $d_{anti-cp}(P(b) \wedge \neg P(a)) = (w_2, \{a, b\})$, while $d_{anti-cp}(P(a) \wedge \neg P(b)) = (\varnothing, \{a, b\})$.

4.4. Discussion

1. In the above analysis, we compared the cp doctrine only with the anti-cp doctrine. Ideally, we would like to define a general class of interpretive doctrines, and establish the cp doctrine as the optimal one among them. This is a task for future research. Here, we briefly remark on what care one should take when pursuing this problem. Consider a constant "doctrine", d, that maps any contractual sentence to the same event with a non-empty reference, say R. Under such a "doctrine", the rich-object assumption will never hold; and, with luck, Z may happen to be the same of R, making perfect insurance possible. Should d be in the feasible set of the optimal doctrine design problem? One may argue not, because d is insensitive to society's interpretations of contractual language, and hence is hardly a legal *interpretive* doctrine. But then what is the appropriate definition for legal interpretive doctrines? This is a question that a full-blown optimal doctrine design exercise needs to address first.
   A reasonable approach would be to define a legal interpretive doctrine as any function d such that $d(\phi) \in \{l_1(\phi), l_2(\phi)\}$ for every $\phi \in \mathscr{L}$. Under this definition, Proposition 7 can be strengthened as follows.

   Proposition 8. *Whenever the Rich-Object Assumption holds under the cp doctrine, it will also hold under any legal interpretive doctrine.*

The proof is the same as that of Proposition 7.

2. Our rationale for the cp doctrine actually does not depend on the assumption that the drafter of the contract has strictly richer awareness than the other party. For example, our argument continues to go through even if agent 2 is also aware of an array of (real) objects that agent 1 is not aware of. Those objects will play no role in the analysis, because the drafter, by definition, cannot write any sentence that makes reference to objects that she is unaware of. Additionally, suppose that there is an array of (real) objects that both agents 1 and 2 are aware of. The rationale behind the cp doctrine seems intuitive enough that it should be robust with respect to this complication as well, although the statements of the Rich-Object Assumption and of Proposition 6 would not be as clean.

3. Our analysis of the slip-the-mind case may seem surprising to the reader, especially in light of the recent literature where various authors have obtained interesting results in insurance contract design when the insured lacks full awareness. Let's point out an implicit assumption that differentiates our work from the rest. We assume that, after agent 2 reads a contract that reminds him of some objects that had previously slipped his mind, he continues to assign the same probability to the event of a negative income shock as before. If this assumption seems implausible, recall that in our framework it is possible for an agent to (explicitly) believe that *something* has slipped his mind, even though he is not aware of anything that has; hence he is not surprised when he later on comes across an example of such a thing. An agent's awareness and his (implicit) beliefs are logically distinct. While one could also tell stories where there is some link between the two, our present aim is to consider what difficulties are imposed on contracting parties by lack of awareness alone. To this end, we work with a model that captures this issue but isolates it from all others. We recognize that a fully-fledged theory of insurance contracts would need to address more systematically the question of how an agent's knowledge, probabilistic beliefs, and awareness change when he is exposed to new information that makes reference to objects that he was unaware of earlier. Developing models that do just this is a priority for our future research.

5. SPECULATIVE TRADE

In this section, we use the OBU structures to study the possibility of speculative trade under unawareness.[17] It is well known that, in classical state-space models with a common prior, common knowledge of strict willingness to trade is impossible when agents are non-delusional (i.e., if they never hold false belief [18]); on the other hand, when agents are delusional, speculative trade may occur. This result remains true when there is unawareness. Here we present two new results that we believe will be of some interest: either if everyone is *living in denial* (i.e., believes, perhaps incorrectly, that there is nothing they are unaware of), or if everyone is *living in paranoia* (i.e., believes, perhaps incorrectly, that there is something that they are unaware of), common knowledge of strict willingness to trade is still impossible, notwithstanding the fact that the agents may be delusional. The proof of this result makes use of an auxiliary theorem which is of interest on its own. The auxiliary theorem states that speculative trade is impossible as long as agents are *terminally partitional*, and hence generalizes the c6lassical no-trade theorem even in standard state-space models.[19]

5.1. Review of the Classical No-Trade Theorem

An OBU structure with common prior is given by $\langle W, O, \{O_w\}, \{\mathscr{I}_i\}, \{\mathscr{A}_i\}, q\rangle$, where W is finite (see Section 3). For the remainder of this section we assume that the information functions $\mathscr{I}_i$ satisfy *belief consistency*, i.e. for all $w \in W$ and all i, $\mathscr{I}_i(w) \neq \varnothing$. Belief consistency guarantees that conditional expectations are well defined. Given any OBU structure with common prior, we shall

[17] Heifetz et al. (2013) also study the possibility of speculative trade under unawareness, in a rather different framework from our own. They do not study situations where agents are living in denial or in paranoia.

[18] So far, we have been talking about what an agent knows and does not know, and interpreting L_i and K_i as *knowledge* operators. But these operators can also be interpreted as representing what an agent believes. Typically, it is assumed that one of the differences between knowledge and belief is that while truth is a necessary condition for knowledge, one may believe in something that is false. Since the main aim of this section is to analyze the implications of various assumptions about what is true, it may be clearer and more appropriate to talk about belief in this section, and be very explicit about truth/falsehood.

[19] Geanakoplos (1989) provides other generalizations of the classical no-trade theorem. The five conditions studied there (nondelusion, knowing that you know, nestedness, balancedness, and positively balancedness) neither imply nor are implied by terminal partitionality.

call the corresponding pair $\langle W, \{\mathscr{I}_i\}\rangle$ its *Kripke frame* (after the logician Saul Kripke).

With two additional restrictions on the information functions, Kripke frames form the basis of the standard model of information used in the economics literature:

- Non-delusion: for all $w \in W$ and all i, $w \in \mathscr{I}_i(w)$.
- Stationarity: for all $w, w' \in W$ and all i, if $w' \in \mathscr{I}_i(w)$ then $\mathscr{I}_i(w) = \mathscr{I}_i(w')$.

We refer to these two assumptions jointly as *partitionality*, since together they imply that $\mathscr{I}_i$ defines a partition on W. A Kripke frame that satisfies non-delusion and stationarity is often referred to as an *Aumann structure* or *information partition model*. Intuitively, non-delusion implies that if an agent (implicitly) believes a fact, then that fact is true; stationarity implies that agents believe that they believe what they actually do believe (positive introspection) and also believe that they don't believe what they actually don't believe (negative introspection).

Let $v : W \to \mathbb{R}^I$ be a function that satisfies $\sum_i v_i(w) = 0$ for every state w. The function v can be thought of as a trade contract that specifies the net monetary transfer to each agent in each state. Let F_i^v denote the event with empty sense (i.e., $sen(F_i^v) = \varnothing$) and with reference equal to the subset of worlds in which agent i's conditional expection of v is strictly positive:

$$ref(F_i^v) = \left\{ w \,\middle|\, \frac{\sum_{w' \in \mathscr{I}_i(w)} q(w')v_i(w')}{\sum_{w' \in \mathscr{I}_i(w)} q(w')} > 0 \right\}.$$

F_i^v can be interpreted as the event that agent i has strict willingness to trade. Let F^v be the conjunction of F_i^v's for every i (i.e., $F^v = \wedge_i F_i^v$), so that F^v is the event that every agent has strict willingness to trade. Let $\mathsf{K}^n F^v$ be recursively defined as $\wedge_i \mathsf{K}_i \mathsf{K}^{n-1} F^v$, with $\mathsf{K}^0 F^v = F^v$. Finally, define

$$\mathsf{CK}F^v := \wedge_{n \geq 1} \mathsf{K}^n F^v.$$

Clearly, $\mathsf{CK}F^v$ is the event that it is a common belief that every agent has strict willingness to trade. The *no-trade* happens if $ref(\mathsf{CK}F^v) = \varnothing$ for every trade contract v. On the other hand, if $w \in \mathsf{CK}F^v$ for some v and w, then *speculative trade* occurs.

The following result is a straightforward translation of the classical no-trade theorem to our setting. See, for example, Samet (1998) for a proof.

Proposition 9. *Take any OBU structure with common prior. If it satisfies non-delusional and stationarity (i.e., if it is partitional), then the no-trade result obtains.*

It is also well known that stationarity alone, without non-delusion, does not suffice to guarantee the no-trade result, nor does non-delusion alone without stationarity. In the next subsection, we prove a stronger version of the classical No-Trade Theorem, which says that the no-trade result still obtains when partitionality is weakened to a condition we call *terminal partitionality*.

5.2. Terminal Partitionality

Given any OBU structure, let $\langle W, \{\mathscr{I}_i\}\rangle$ be its Kripke frame. We first generalize the notion of partitionality to subspaces of W: $W' \subseteq W$ is partitional if for all $w, w' \in W'$, $\mathscr{I}_i(w) \subseteq W'$ for all i, and also non-delusion and stationarity are satisfied. Next, for every subspace $W' \subseteq W$, define

$$D(W') = \{w \in W \mid \mathscr{I}_i(w) \subseteq W' \text{ for some agent } i\}.$$

$D(W')$ is the collection of worlds in which at least one agent considers only worlds in W' to be possible. We say that an OBU structure (and its Kripke frame) satisfies *terminal partitionality* if there is a non-empty partitional subspace $W' \subseteq W$ such that $\cup_{n\geq 0} D^n(W') = W$, where $D^n(W')$ is defined recursively as $D(D^{n-1}(W'))$, and $D^0(W') = W'$.

Note that terminal partitionaity is a strictly weaker condition than partitionality. It says that there is a subset of states where agents satisfy non-delusion and stationarity (i.e. where everything they believe is true and they have access to their own beliefs), and in every other state, some agent either believes that everyone satisfies non-delusion and stationarity, or believes that someone believes that everyone satisfies non-delusion and stationarity, or believes that someone believes that someone believes that ....

The next proposition says that the condition of partitionality in the classical no-trade theorem can be replaced by terminal partitionality.

Proposition 10. *Take any OBU structure with common prior. If it is terminally partitional, then the no-trade result obtains.*

Proof. Let $\langle W, \{\mathscr{I}_i\}\rangle$ be the corresponding Kripke frame, and let W' be a partitional subspace such that $\cup_{n\geq 0} D^n(W') = W$. Such a partitional subspace exists by assumption. We prove by induction that

$$ref(\mathsf{CK}F^v) \cap D^n(W') = \varnothing \tag{10}$$

for every n, which implies that $ref(\mathsf{CK}F^v) = ref(\mathsf{CK}F^v) \cap W = \varnothing$, completing the proof. For $n = 0$, this follows from Proposition 9 (applied to the sub-structure with state space W').

For the inductive step, suppose equation (10) has been proved up to n; we prove it for $n+1$. Consider any world $w \in D^{n+1}(W')$; i.e., any world w such that $\mathscr{I}_i(w) \subseteq D^n(W')$ for some agent i. Suppose $w \in ref(\mathsf{CK}F^v)$. Then $w \in ref(\mathsf{K}_i\mathsf{K}^m F^v)$ for every $m \geq 1$, and hence $\mathscr{I}_i(w) \subseteq ref(\mathsf{K}^m F^v)$ for every $m \geq 1$. Therefore $\mathscr{I}_i(w) \subseteq ref(\mathsf{CK}F^v)$. But then $ref(\mathsf{CK}F^v) \cap D^n(W') \supseteq \mathscr{I}_i(w) \neq \varnothing$ yields a contradiction. So we have $ref(\mathsf{CK}F^v) \cap D^{n+1}(W') = \varnothing$, as required. □

5.3. Living in Denial and Living in Paranoia

Informally, we say that an agent is *living in denial* if she always believes that there is nothing she is unaware of (although there may be). Similarly, we say that she is *living in paranoia* if she always believes that there is something she is unaware of (although there may be none). Let's illustrate these two concepts with two examples before getting into the formality.

Example 1. *Consider an OBU structure with only one agent;* $W = \{w_1, w_2\}$; $O = \{o_1, o_2\}$, $O_{w_1} = \{o_1\}$, $O_{w_2} = \{o_1, o_2\}$; $\mathscr{A}(w_1) = \mathscr{A}(w_2) = \{o_1\}$; *and* $\mathscr{I}(w_1) = \mathscr{I}(w_2) = \{w_1\}$.

In this example, although the agent is aware of exactly the same object in both states (i.e., $\mathscr{A}(w_1) = \mathscr{A}(w_2)$), different things are true in these states. In particular, in w_1 there is nothing that the agent is unaware of, while in w_2 there is something that the agent is unaware of. Note that in both states, the agent considers only w_1 as possible. Therefore the agent is delusional in w_2: she believes that there is nothing she is unaware of when there actually is. In this example, the agent always believes that there is nothing she is unaware of (although there may be), and hence she is living in denial.

Example 2. *Consider an OBU structure which is the same as in Example 1, except for that the information function is now* $\mathscr{I}(w_1) = \mathscr{I}(w_2) = \{w_2\}$.

In this example, in both states w_1 and w_2, the agent considers only w_2 possible. Therefore the agent is delusional in world w_1: she believes that there is something she is unaware of when there actually is none. In this example, the agent always believes that there is something she is unaware of (although there may be none), and hence she is living in paranoia.

Of course there is no reason why agents who are living in denial could not coexist with agents who are living in paranoia. An interesting task for future research is to study strategic interaction among these different kinds of agents. For now, however, we focus on cases where everyone is living in denial, or where everyone is living in paranoia.

Note that an agent who is living in denial may be delusional, and the classical no-trade theorem (Proposition 9) does not rule out the possibility of speculative trade. But living in denial, when it gives rise to delusion, results in a very specific form of delusion. In fact, we show that if this is the only form of delusion suffered by the agents, then speculative trade is still impossible. A similar result holds for the case where everyone is living in paranoia.

Definition 11. *An OBU structure satisfies WLID (weak living-in-denial) if, for every state w and agent i,*

1. $\mathscr{A}_i(w) \subseteq O_w$;
2. $\mathscr{A}_i(w') = O_{w'}$ *for every* $w' \in \mathscr{I}_i(w)$; *and*
3. $\mathscr{A}_i(w) = O_w$ *implies* $w \in \mathscr{I}_i(w)$ *and* $\mathscr{I}_i(w') = \mathscr{I}_i(w)$ *for* $w' \in \mathscr{I}_i(w)$.

The second part of the definition says that agent i considers possible only states in which she is aware of everything, and so she believes (correctly or incorrectly) that there is nothing she is unaware of. The third part says that if this belief turns out to be correct in a given state, then she has no false beliefs in that state and has access to her own beliefs.

Definition 12. *An OBU structure satisfies WLIP (weak living-in-paranoia) if, for every state w and agent i,*

1. $\mathscr{A}_i(w) \subseteq O_w$;
2. $\mathscr{A}_i(w') \subsetneq O_{w'}$ *for every* $w' \in \mathscr{I}_i(w)$; *and*
3. $\mathscr{A}_i(w) \subsetneq O_w$ *implies* $w \in \mathscr{I}_i(w)$ *and* $\mathscr{I}_i(w') = \mathscr{I}_i(w)$ *for* $w' \in \mathscr{I}_i(w)$.

WLIP is the opposite of WLID in some sense: every agent believes (correctly or incorrectly) that there is something she is unaware of; and if she turns out to be correct about this, she is correct on every other matter and also has access to her own beliefs.

Both WLID and WLIP are "weak" conditions in the sense that even a partitional OBU structure can satisfy WLID or WLIP (although it cannot satisfy both simultaneously).

Before we state our main results, we need one more definition. We say that an OBU structure satisfies *LA-introspection* if, for every state w and every agent i, $w' \in \mathscr{I}_i(w)$ implies $\mathscr{A}_i(w') = \mathscr{A}_i(w)$. LA-introspection is characterized by Board & Chung (2021)'s axioms **LA1** and **LA2**, which jointly say that every agent has correct beliefs about what she is aware of (see Board & Chung (2021) for more details).

Proposition 13. *Consider an OBU structure with common prior, and suppose it satisfies WLID and LA-introspection. Then it also satisfies terminal partitionality.*

Proof. For any two worlds, w and w', we say that w *points to* w' if there is an agent i such that $w \notin \mathscr{I}_i(w)$ and $w' \in \mathscr{I}_i(w)$.

Suppose w points to w'. Then $w \notin \mathscr{I}_i(w)$ for some agent i. By WLID, LA-introspection, and WLID again, we have

$$O_{w'} = \mathscr{A}_i(w') = \mathscr{A}_i(w) \subsetneq O_w \tag{11}$$

for some agent i. Therefore a world can only point to other worlds that have strictly smaller sets of real objects. Then, by finiteness of W, there exist worlds that do not point to any other worlds. Let W' be the collection of these worlds.

If w belongs to W', then $w \in \mathscr{I}_i(w)$ for any agent i. Furthermore, for any agent i, by the second and the third parts of WLID respectively, we have $\mathscr{A}_i(w) = O_w$ and hence $\mathscr{I}_i(w') = \mathscr{I}_i(w)$ for any $w' \in \mathscr{I}_i(w)$. But this means $w' \in \mathscr{I}_i(w)$ implies $w' \in \mathscr{I}_i(w')$, and hence w' also does not point to any other worlds. Therefore W' is a partitional subspace.

If $W \neq W'$, then by finiteness of $W \setminus W'$, and by the observation that a world can only point to worlds that have strictly smaller sets of real objects, there must exist worlds in $W \setminus W'$ that do not point to any other worlds in $W \setminus W'$. Let W'' be the collection of these worlds. It is easy to verify that

$D(W') = W'' \cup W' \supsetneq W'$. Repeating this argument, one can show that if $W \neq D^n(W')$, then $D^{n+1}(W')$ is a strict superset of $D^n(W')$. Therefore, by finiteness again, $W = \cup_{n \geq 0} D^n(W')$. □

Proposition 14. *Consider an OBU structure with common prior, and suppose it satisfies WLIP and LA-introspection. Then it also satisfies terminal partitionality.*

Proof. The proof is similar to that of Proposition 13, except for equation (11). Suppose w points to w'. Then $w \notin \mathscr{I}_i(w)$ for some agent i. By WLIP, LA-introspection, and WLIP again, we have

$$O_{w'} \supsetneq \mathscr{A}_i(w') = \mathscr{A}_i(w) = O_w$$

for some agent i. Therefore a world can only point to other worlds that have strictly larger sets of real objects. The rest of the proof now follows the same arguments as in that of Proposition 13. □

Corollary 15. *Consider a regular OBU structure with common prior, and suppose it satisfies LA-introspection. If it satisfies either WLID or WLIP, then the no-trade result obtains.*

Proof. This follows from Propositions 10, 13 and 14. □

6. CONCLUSION

As we discussed in the introduction, there is large gap in the literature on unawareness between the more applied studies that appeal to unawareness to motivate the assumptions underlying their models, and the foundational studies that often pay little attention to the real-world applications. In this and our companion papers, we have attempted to bridge this gap. In particular, we have shown that the key assumption in the applied literature, namely that agents are "unaware, but know that they are unaware", can be captured in a rational-agent framework; furthermore, this assumption is perfectly consistent with the DLR axioms that much of the foundational literature tries to accommodate.

Although the OBU structures described above derive an agent's unawareness of propositions from her unawareness of the objects mentioned in those propositions, one can envisage an extension where unawareness of properties is also modeled. A property-unawareness function could work (roughly) as follows: if an agent is unaware of a given property, then she would be unaware of any event containing one state but not another, where the two states could only be distinguished by whether or not various objects satisfied that property. Combining such a property-unawareness function with the object-unawareness function analyzed above would allow us to separate two kinds of unawareness: and agent could be unaware that "Yao Ming is tall" either because she has no idea who Yao Ming is or because she does not understand the concept of height.

In additional to providing foundations for a model of unawareness, in the form of OBU structures, we have also presented two applications: the first examines the legal interpretive doctrine contra proferentem, while the second extends the classical no trade theorem to cover cases where agents are mistaken in a particular away (they live in denial or in paranoia). These applications, we hope, will convince the reader that it is straightforward to use OBU structures in applied work. We also believe that the results of these applications are valuable in their own right.

Before we finish, we would also like to mention a recent experimental paper that provides evidence suggesting that agents may be unsure whether they are aware of everything or not. Blume & Gneezy (2010) have their subjects play a game with each other. There is a less-obvious strategy that guarantees a win, and a more-obvious strategy that results in a win half the time. Even though a win paid out $10, some subjects rejected an outside option of $6 and then played the more-obvious strategy, for an expected payout of $5. Presumably these subjects were not aware of the less-obvious strategy. Why then did they reject the outside option? Blume & Gneezy (2010) suggest that this is because they believed such a strategy existed, and hoped to figure it out after rejecting the outside option but before playing the game. In our language, we would say that these agents believed there was something they were unaware of, and hoped to become aware of it in the future.

Appendix A: Model-Theoretic Description of OBU Structures

For the sake of transparency, and to aid interpretation, we now show how OBU structures assign truth conditions for a formal language, a version of first-order modal logic.[20] We start with a set of (unary) predicates, $P, Q, R, \ldots$, and an (infinite) set of variables, $x, y, z, \ldots$. Together with set of objects, O, this generates a set Φ of atomic formulas, $P(\mathsf{a}), P(x), Q(\mathsf{a}), Q(x), \ldots$, where each predicate takes as its argument a single object or variable. Let $\mathscr{F}$ be the smallest set of formulas that satisfies the following conditions:

- if $\phi \in \Phi$, then $\phi \in \mathscr{F}$;
- if $\phi, \psi \in \mathscr{F}$, then $\neg\phi \in \mathscr{F}$ and $\phi \wedge \psi \in \mathscr{F}$;
- if $\phi \in \mathscr{F}$ and $x \in X$, then $\forall x \phi \in \mathscr{F}$;
- if $\phi \in \mathscr{F}$, then $\mathtt{L}_i\phi \in \mathscr{F}$ and $\mathtt{A}_i\alpha \in \mathscr{F}$ and $\mathtt{K}_i\alpha \in \mathscr{F}$ for each agent i.

Formulas should be read in the obvious way; for instance, $\forall x \mathtt{A}_i P(x)$ is to be read as "for every x, agent i is aware that x possesses property P." Notice, however, that it is hard to make sense of certain formulas: consider $P(x)$ as opposed to $P(a)$ or $\forall x P(x)$. Although it may be reasonable to claim that a specific object, a, is P, or that every x is P, the claim that x is P seems empty unless we specify which object variable x stands for. In general, we say that a variable x is free in a formula if it does fall under the scope of a quantifier $\forall x$, and define our language $\mathscr{L}$ to be the set of all formulas containing no free variables.[21] We use OBU structures to provide truth conditions only for formulas in $\mathscr{L}$, and not for formulas such as $P(x)$ that contain free variables.

Take an OBU structure $M = \langle W, O, \{O_w\}, \{\mathscr{I}_i\}, \{\mathscr{A}_i\}\rangle$, and augment it with an *assignment* $\pi(w)(P) \subseteq O$ of objects to every predicate at every state (intuitively, $\pi(w)(P)$ is the set of objects that satisfy predicate P). If a formula $\phi \in \mathscr{L}$ is true at state w of OBU structure M under assignment π, we write $(M, w, \pi) \vDash P(\mathsf{a})$; $\vDash$ is defined inductively as follows:

[20] Board & Chung (2021) provide the (model-theoretic) soundness and complete axiomatization.

[21] More formally, we define inductively what it is for a variable to be free in $\phi \in \mathscr{F}$:

- if ϕ is an atomic formula of the form $P(x)$ where x is a variable, then x is free in ϕ;
- x is free in $\neg\phi$, $\mathtt{K}_i\phi$, $\mathtt{A}_i\phi$, and $\mathtt{L}_i\phi$ iff x is free in ϕ;
- x is free in $\phi \wedge \psi$ iff x is free in ϕ or ψ;
- x is free in $\forall y \alpha$ iff x is free in ϕ and x is different from y.

$(M,w,\pi) \vDash P(\mathsf{a})$ iff $\mathsf{a} \in \pi(w)(P)$;

$(M,w,\pi) \vDash \neg\phi$ iff $(M,w,\pi) \nvDash \phi$;

$(M,w,\pi) \vDash \phi \wedge \psi$ iff $(M,w,\pi) \vDash \phi$ and $(M,w,\pi) \vDash \psi$;

$(M,w,\pi) \vDash \forall x\phi$ iff $(M,w,\pi) \vDash \phi[\mathsf{a}\backslash x]$ for every $a \in O_w$ (where $\phi[\mathsf{a}\backslash x]$ is ϕ with all free occurrences of x replaced with a);

$(M,w,\pi) \vDash \mathtt{A}_i\phi$ iff $\mathsf{a} \in \mathscr{A}_i(w)$ for every object a in ϕ;

$(M,w,\pi) \vDash \mathtt{L}_i\phi$ iff $(M,w',\pi) \vDash \phi$ for all $w' \in \mathscr{I}_i(w)$;

$(M,w,\pi) \vDash \mathtt{K}_i\phi$ iff $(M,w,\pi) \vDash \mathtt{A}_i\phi$ and $(M,w,\pi) \vDash \mathtt{L}_i\phi$.

Notice that there is a close connection between sentences of $\mathscr{L}$ and OBU events: for any given $\phi \in \mathscr{L}$, the reference of the corresponding OBU event is given by the set of states at which ϕ is true, while the sense is simply the set of objects in ϕ.

Appendix B: Proofs

Proof of Proposition 1. 1. Straightforward.

2. Take some A_i' which satisfies A1–A4, and define $\mathscr{A}_i$ as follows: $a \in \mathscr{A}_i(w)$ iff $w \in ref\,\mathsf{A}_i'(W,a)$. We need to show that $\mathsf{A}_i'(R,S) = \mathsf{A}_i(R,S)$. We consider two cases:
Case 1: $S \neq \varnothing$. Then

$$\begin{aligned}
\mathsf{A}_i'(R,S) &= \mathsf{A}_i'(W,S) \text{ (by A2)} \\
&= \wedge_{a\in S}\mathsf{A}_i'(W,a) \text{ (by A1)} \\
&= \wedge_{a\in S}(\{w \mid x \in \mathscr{A}_i(w)\},a) \text{ (by A4 and the definition of } \mathscr{A}_i) \\
&= (\{w \mid S \subseteq \mathscr{A}_i(w)\},S) \text{ (definition of } \wedge) \\
&= \mathsf{A}_i(R,S), \text{ as required.}
\end{aligned}$$

Case 2: $S = \varnothing$. Then

$$\begin{aligned}
\mathsf{A}_i'(R,\varnothing) &= (W,\varnothing) \text{ (by A3)} \\
&= (\{w \in W \mid \varnothing \subseteq \mathscr{A}_i(w)\},\varnothing) \\
&= \mathsf{A}_i(R,\varnothing), \text{ as required.}
\end{aligned}$$

□

Proof of Proposition 2. 1. Straightforward.

2. Take some L'_i which satisfies L1–L4, and define $\mathscr{I}_i$ as follows:

$$\mathscr{I}_i(w) = \left\{w' \mid w \in ref\left(\neg \mathsf{L}'_i \neg\left(w', O\right)\right)\right\}.$$

Note that, by L4,

$$\left\{w' \mid w \in ref\left(\neg \mathsf{L}'_i \neg\left(w', O\right)\right)\right\} = \left\{w' \mid w \in ref\left(\neg \mathsf{L}'_i \neg\left(w', S\right)\right)\right\}$$

for all $S \subseteq O$, so $w' \in \mathscr{I}_i(w)$ iff $w \in ref(\neg \mathsf{L}'_i \neg(w', S))$, and hence

$$w' \notin \mathscr{I}_i(w) \text{ iff } w \in ref\left(\mathsf{L}'_i \neg\left(w', S\right)\right). \tag{$*$}$$

We need to show that $\mathsf{L}'_i(R,S) = \mathsf{L}_i(R,S)$. We consider two cases:
Case 1: $R \neq W$. Then

$$\begin{aligned}
\mathsf{L}'_i(R,S) &= \mathsf{L}'_i\left(\cap_{w \notin R} W \setminus \{w\}, S\right) \\
&= \wedge_{w \notin R} \mathsf{L}'_i(W \setminus \{w\}, S) \text{ (by L2)} \\
&= \wedge_{w \notin R} \mathsf{L}'_i \neg(w, S) \text{ (definition of } \neg) \\
&= \wedge_{w \notin R}\left(\left\{w' \mid w \notin \mathscr{I}_i\left(w'\right)\right\}, S\right) \text{ (by } (*) \text{ and L3)} \\
&= \left(\cap_{w \notin R}\left\{w' \mid w_1 \notin \mathscr{I}_i\left(w'\right)\right\}, S\right) \text{ (definition of } \wedge) \\
&= \left(\left\{w' \mid \mathscr{I}_i\left(w'\right) \subseteq R\right\}, S\right) \\
&= \mathsf{L}_i(R,S) \text{, as required.}
\end{aligned}$$

Case 2: $R = W$. Then $\mathsf{L}'_i(W,O) = (W,O)$ (by L1), so $\mathsf{L}'_i(W,S) = (W,S)$ (by L4). And $\mathsf{L}_i(W,S) = (\{w \mid \mathscr{I}_i(w) \subseteq W\}, S) = (W,S)$.

□

Proof of Proposition 3. 1. Straightforward.

2. Take some All' which satisfies All1–All4. For any $w \in W$ and $a \in O$, construct the property p_{wa} such that:

$$p_{wa}(b) = \begin{cases} (W, b) & \text{if } b \neq a \\ (W \setminus \{w\}, b) & \text{if } b = a \end{cases}.$$

Observe for later use that, by All2, $W \setminus \{w\} \subseteq ref(\mathsf{All}'\, p_{wa})$, and hence, for any $R \subseteq W$,

$$\cap_{w \notin R}\, ref(\mathsf{All}'\, p_{wa}) = \{w | w \in ref(\mathsf{All}'\, p_{wa})\} \cup R. \tag{12}$$

We define $\{O_w\}_{w \in W}$ using these p_{wa}'s as follows:

$$O_w = \left\{a \mid w \notin ref\left(\mathsf{All}'\, p_{wa}\right)\right\}.$$

These O_w's define the property *re*:

$$R_a^{re} = \{w \mid w \notin ref(\mathsf{All}'\, p_{wa})\}.$$

This property *re*, of course, in turn defines the operator All. We need to show that $\mathsf{All}' = \mathsf{All}$. Take an arbitrary property $\tilde{p}$. From All4, we have $sen(\mathsf{All}'\, \tilde{p}) = S^{\tilde{p}}$; and $sen(\mathsf{All}\, \tilde{p}) = S^{\tilde{p}}$ from the definition of All. It remains to show that $ref(\mathsf{All}'\, \tilde{p}) = ref(\mathsf{All}\, \tilde{p})$.
From $\tilde{p}$, construct another property $\hat{p}$ as follows:

$$\hat{p} := \wedge_{a \in O} \wedge_{w \notin R^{\tilde{p}a}}\ p_{wa}.$$

We claim that $R_b^{\hat{p}} = R_b^{\tilde{p}}$ for every $b \in O$, and hence by All3, we have $ref(\mathsf{All}'\, \hat{p}) = ref(\mathsf{All}'\, \tilde{p})$. To prove this claim, notice that, for any $b \in O$,

$$\begin{aligned} R_b^{\hat{p}} &= \cap_{a \in O} \cap_{w \notin R_a^{\tilde{p}}} R_b^{p_{wa}} \\ &= \Big(\cap_{\substack{a \neq b \\ w \notin R_a^{\tilde{p}}}} R_b^{p_{wa}}\Big) \cap \Big(\cap_{\substack{a = b \\ w \notin R_a^{\tilde{p}}}} R_b^{p_{wa}}\Big) \\ &= \Big(\cap_{\substack{a \neq b \\ w \notin R_a^{\tilde{p}}}} W\Big) \cap \Big(\cap_{w \notin R_b^{\tilde{p}}} W \setminus \{w\}\Big) \\ &= R_b^{\tilde{p}}, \text{ as required.} \end{aligned}$$

Therefore, it suffices to prove that $ref(\mathsf{All}'\,\hat{p}) = ref(\mathsf{All}\,\tilde{p})$. By All1, we have

$$\begin{aligned}
ref(\mathsf{All}'\,\hat{p}) &= \cap_{a\in O} \cap_{w\notin R_a^{\tilde{p}}} ref(\mathsf{All}'\,p_{wa}) \\
&= \cap_{a\in O} \left(\{w \mid w \in ref(\mathsf{All}'\,p_{wa})\} \cup R_a^{\tilde{p}}\right) \quad \text{(by (12))} \\
&= \cap_{a\in O} \left(R_a^{\neg re} \cup R_a^{\tilde{p}}\right) \\
&= \cap_{a\in O} R_a^{\neg re \vee \tilde{p}} \\
&= \cap_{a\in O} R_a^{re \rightarrow \tilde{p}} \\
&= ref(\mathsf{All}\,\tilde{p}), \text{ as required.}
\end{aligned}$$

□

References

Abraham, K. S. (1996). A theory of insurance policy interpretation. *Michigan Law Review*, *95*(3), 531–569.

Blume, A., & Gneezy, U. (2010). Cognitive forward induction and coordination without common knowledge: An experimental study. *Games and Economic Behavior*, *68*(2), 488-511.

Board, O. J., & Chung, K.-S. (2021). Object-based unawareness: Axioms. *Journal of Mechanism and Institution Design*, *6*(6), 1–36.

Carolina Care Plan Incorporated v. McKenzie. (2007). *551 U.S. 1* (No. 06-1182 (R46-022)).

Chung, K.-S., & Fortnow, L. (2016). Loopholes. *The Economic Journal*, *126*(595), 1774-1797.

Dekel, E., Lipman, B. L., & Rustichini, A. (1998). Standard state-space models preclude unawareness. *Econometrica*, *66*(1), 159–173.

Fagin, R., & Halpern, J. Y. (1987). Belief, awareness, and limited reasoning. *Artificial Intelligence*, *34*(1), 39-76.

Filiz-Ozbay, E. (2012). Incorporating unawareness into contract theory. *Games and Economic Behavior*, *76*(1), 181-194.

Geanakoplos, J. (1989). Game Theory Without Partitions, and Applications to Speculation and Consensus. *Cowles Foundation Discussion Papers*, *914*.

Gul, F. (1998). A comment on Aumann's Bayesian view. *Econometrica*, *66*(4), 923–927.

Halpern, J. Y. (1999). Set-theoretic completeness for epistemic and conditional logic. *Annals of Mathematics and Artificial Intelligence*, *26*(1), 1-27.

Halpern, J. Y. (2001). Alternative semantics for unawareness. *Games and Economic Behavior*, *37*(2), 321-339.

Halpern, J. Y., & Rêgo, L. C. (2006). Reasoning about knowledge of unawareness. In *Proceedings of the Tenth International Conference on Principles of Knowledge Representation and Reasoning* (pp. 6–13). AAAI Press.

Heifetz, A., Meier, M., & Schipper, B. C. (2006). Interactive unawareness. *Journal of Economic Theory*, *130*(1), 78-94.

Heifetz, A., Meier, M., & Schipper, B. C. (2013). Unawareness, beliefs, and speculative trade. *Games and Economic Behavior*, *77*(1), 100-121.

Li, J. (2009). Information structures with unawareness. *Journal of Economic Theory*, *144*(3), 977-993.

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

Modica, S., & Rustichini, A. (1999). Unawareness and partitional information structures. *Games and Economic Behavior*, *27*(2), 265-298.

Morris, S. (1995). The common prior assumption in economic theory. *Economics and Philosophy*, *11*(2), 227–253. doi: 10.1017/S0266267100003382

Ozbay, E. Y. (2007). Unawareness and strategic announcements in games with uncertainty. In *Proceedings of the 11th Conference on Theoretical Aspects of Rationality and Knowledge* (p. 231–238). New York, NY, USA: Association for Computing Machinery.

Samet, D. (1998). Common priors and separation of convex sets. *Games and Economic Behavior*, *24*(1), 172-174.

Sillari, G. (2006). Models of awareness. In *Proceedings of the 7th Conference on Logic and the Foundations of Game and Decision Theory (loft)* (pp. 209–218).

Tirole, J. (2009). Cognition and incomplete contracts. *American Economic Review*, *99*(1), 265-94.