

# Partial Awareness

Joseph Halpern and Evan Piermont

## Abstract

We develop a modal logic to capture partial awareness. The logic has three building blocks: objects, properties, and concepts. Properties are unary predicates on objects; concepts are Boolean combinations of properties. An agent can then be partially aware of a concept if she is aware of the concept in the abstract without being aware of the properties that define it. The logic also allows for quantification over objects and properties so that the agent can reason about her own unawareness. We then apply the logic to contracts. Contracts are syntactic objects that dictate outcomes based on the truth of formulae. The set of feasible contracts is limited by the formulae the agents are aware of. We show that when agents are unaware, referencing concepts that agents are only partially aware of can improve welfare.

## 1 Introduction

Standard models of epistemic logic assume that agents are *logically omniscient*: they know all valid formulas and logical consequences of their knowledge. There have been many attempts to find models of knowledge that do not satisfy logical omniscience. One of the most common approaches involves *awareness*. Roughly speaking, an agent  $i$  cannot know a valid formula  $\varphi$  if  $i$  is unaware of  $\varphi$ . For example, an agent cannot know that either quantum computers are faster than conventional computers or they are not if she is not aware of the notion of quantum computer.

There have been many attempts to capture unawareness in the computer science, economics, and philosophy literature, ranging from syntactic approaches (Fagin and Halpern 1988), to semantic approaches involving lattices (Heifetz, Meier, and Schipper 2006), to identifying the lack of awareness of  $\varphi$  with an agent neither knowing  $\varphi$  nor knowing that she does not know  $\varphi$  (Modica and Rustichini 1994; 1999). Most of the attempts involved propositional (modal) logics, although there are papers that use first-order quantification as well (Board and Chung 2009; Sillari 2008). However, none of these approaches are rich enough to capture what we will call *partial unawareness*.

Perhaps the most common interpretation of lack of awareness identifies the lack of awareness of  $\varphi$  with the sentiment “ $\varphi$  is not on my radar screen”. With this interpreta-

tion, partial awareness becomes “some aspects of  $\varphi$  are on my radar screen”.<sup>1</sup> Consider an agent who is in the market for a new computer. She might be completely unaware of quantum computers, never having heard of one at all. Such an agent cannot reason about her value for having a quantum computer, nor think about for which tasks a quantum computer would be useful. But this is a rather extreme case. A slightly more aware agent might be aware of (the concept of) a quantum computer, having recently read a magazine article on the topic. She might understand some properties of quantum computers, for example, that they can factor integers faster than a conventional computer, but be unaware of the notion of qubit state on which quantum computing is based. Such an agent may well be able to reason about her value of a quantum computer despite her less than full awareness.

To capture such partial awareness more formally, we consider a logic with three building blocks: *objects*, *properties*, and *concepts*. We take a property to be a unary predicate, so it denotes a subset of objects in the domain;<sup>2</sup> a concept is a Boolean combination of properties. In each state (possible world), each agent is aware of a subset of objects, properties, and concepts.

The use of concepts in the context of awareness, which is (to the best of our knowledge) original to this paper, is critical in our approach, and is how we capture *partial* awareness. For a simple example of how we use it, suppose that a quantum computer ( $Q$ ) is defined as a computer ( $C$ ) that possess an additional “quantum property”  $QP$ . That is,  $Q$  is defined to be  $C \wedge QP$  (more precisely, we will have  $\forall x(Q(x) \Leftrightarrow C(x) \wedge QP(x))$  as a domain axiom). A “partially aware” agent might be aware of the concept of a quantum computer but unaware of the specific Boolean combination of properties that characterizes it. Contrast this to the cases where the agent is fully unaware (she is unaware of

<sup>1</sup>Lack of awareness of  $\varphi$  has also been identified with the inability to compute whether  $\varphi$  is true (due to computational limitations). We do not consider this interpretation in this paper, but partial awareness makes sense for it as well—now partial awareness becomes “I can compute whether some aspects of  $\varphi$  are true.”

<sup>2</sup>We could easily extend our approach to allow arbitrary  $k$ -ary predicates, but this would complicate the presentation. Restricting to unary predicates allows us to focus on the more interesting new conceptual issues.

even the concept of a quantum computer) or fully aware (she is aware of both the concept of a quantum computer and also what it means to be one, i.e., the properties  $C$  and  $QP$ ).

Once we have awareness in the language, we need to consider what an agent knows about her own awareness (and lack of it). This is critical in order to capture many interesting economic behaviors. For example, an agent might know (or at least believe it is possible) that there are aspects of a quantum computer about which she is unaware (in our example, this happens to be  $QP$ ). In the spirit of Halpern and Rêgo (2009; 2013) (HR from now on), we capture this using quantification over properties:  $K_i(\exists P \forall x (Q(x) \Leftrightarrow C(x) \wedge P(x)))$ ; although  $i$  is unaware of *how* quantum computers differ from conventional computers, she knows that there is a distinction that is captured by some property  $P$ .

As models of unawareness become richer, they begin to resemble models of uncertainty. While the agent is unaware of the property  $QP$ , so cannot reason explicitly about it, she is aware that there is some property that relates  $C$  and  $Q$ . This allows her to reason at a sophisticated level about  $QP$ . For example, the statement  $K_i(\forall P \forall x ((Q(x) \Leftrightarrow C(x) \wedge P(x)) \Rightarrow (P(x) \Rightarrow R(x))))$ , combined with the definition of quantum computer, implies that the agent knows that  $QP$  implies  $R$ , even though she is unaware of  $QP$ . Despite her (partial) unawareness of the notion of quantum computer, the agent can reach some substantive conclusions.

With unawareness, an agent may in general be uncertain about the relation between  $C$  and  $Q$ ; for example, she might also envision a state where a quantum computer is a computer that satisfies one of two properties, either  $QP$  or  $QP'$ , but cannot articulate statements that distinguish  $QP$  from  $QP'$ . So if the agent wishes to purchase a  $QP$ -computer but not a  $QP'$ -computer, she cannot do so. This lack of awareness has important consequences in market settings.

Consider a seller of quantum computers, who is fully aware, and who has the ability to teach the buyer—he can expand the buyer’s awareness, allowing her to discriminate between  $QP$  and  $QP'$  computers. It is instructive to compare the case of unawareness with the more standard case of uncertainty (with full awareness) in this setting. In environments of pure uncertainty, the buyer and seller are assumed to have a common (and fully understood) space of uncertainty, where each state fully resolves all payoff-relevant uncertainty. That is, although they may have uncertainty, both the buyer and seller understand exactly what information is required in order to resolve the uncertainty. A state contains all the relevant information, so, given a state, both the buyer and seller can (at least in principle) place a price on all the relevant options.

Suppose that we model the example above using two states:  $s$ , in which  $c$  is a  $QP$ -computer, and  $s'$ , in which  $c$  is  $QP'$ -computer. If the buyer knows that the seller knows the state, and the seller can credibly reveal his information, then we can assume without loss of generality that he will always do so and a transaction will take place only in state  $s$ . To see why, note that whenever the state is  $s$ , the seller’s dominant strategy is to reveal his information, ensuring a sale. Because the buyer knows that the seller knows the state, she will interpret no information as a signal that the state is  $s'$ .

Thus, in either case the state is revealed.<sup>3</sup>

If the buyer does not know whether the seller knows the state, or if the seller cannot credibly reveal the state, the argument above fails; receiving no information could plausibly be the consequence of an uninformed seller. If and when information is revealed or a transaction takes place now depends on the beliefs of the agents. However, an enforceable contract can remedy the situation: a contract that stipulates the sale of the computer conditional on the true state being  $s$  results in essentially the same outcome as the case where the buyer knows that the seller knows the true state of the world: the buyer ends up with the computer if and only if the state is  $s$ . The fact that the buyer and seller agree on the underlying state space (i.e., the set of possible states of the world) makes it possible for an enforceable contract to overcome information asymmetries.

We now turn to the situation with unawareness. If the buyer knows that the seller is aware of all the intricacies of his product, then the analogous argument to the one above shows that we get the same outcome as in the case of uncertainty. That is, if the buyer believes the seller is himself aware of the distinction between  $QP$  and  $QP'$ , then she could decide to purchase a computer  $c$  only if he teaches her about the properties and if  $QP(c)$  is true.

On the other hand, if the buyer does not know what the seller is aware of, then we get a significant divergence between the situation with awareness and uncertainty. With unawareness, the seller will again not volunteer information, but the buyer *cannot* draw up a contract guaranteeing her the product she wants, since she cannot articulate the difference between the states. The efficacy of contracts relies critically on the parties’ common knowledge of the state space, which in general does not hold in the presence of unawareness.

Note the critical role of the “partialness” of awareness here. If the buyer were fully aware of the concept, in the sense of her being aware of the properties that define it, she could write the relevant contracts and we would have a case of pure uncertainty. On the other hand, if she were completely unaware of quantum computers, she would not be able to reason about her value, or even consider buying one.

The introduction of concepts allows us to consider agents with different levels of awareness. For example, perhaps a buyer becomes aware of a particular company that is offering a commercial quantum computer understood to be characterized by the concept  $T$ . If the buyer knows that  $T$  is such that  $\forall x (T(x) \Leftrightarrow C(x) \wedge QP'(x))$ , then the buyer, without being explicitly aware of  $QP$  or  $QP'$ , can nonetheless articulate her desire to purchase a  $QP$  but not a  $QP'$  computer. Indeed, the buyer can write a contract that gives her the right

<sup>3</sup>This argument does not rely on the fact that there are only two states. Suppose that there are  $n$  states, say  $s_1, \dots, s_n$ , and the buyer is willing to pay  $p_i$  for the computer if the true state is  $s_i$ , with  $p_1 \geq p_2 \geq \dots \geq p_n$ . Assume that the seller is willing to sell at any positive price. An easy induction on  $k$  shows that, without loss of generality, if the true state is  $s_k$ , the seller might as well reveal this fact, as long as the buyer puts positive probability on a state  $k' < k$ . (A formalization of this argument also requires common knowledge of rationality, or, more precisely, sufficiently deep knowledge of rationality.)

to return the computer  $c$  in the event that  $T(c)$  is true. In other words, the concept  $T$  acts as a proxy for the property  $QP$ , allowing the buyer to circumvent her scant awareness.

Note that the observations above help to explain the prevalence of costly litigation and contractual disputes. In the world of pure uncertainty, it can be shown that contracts are always upheld in equilibrium. Indeed, if uncertainty resolved in such a way as to make some party renege, then this could be foreseen, and could be addressed by an appropriate contract, avoiding costly litigation. However, when parties are aware of different concepts, optimal complete contracts cannot be drawn up, setting up a barrier to efficient trade. A legal system that punishes the strategic concealment of information can help to facilitate trade, as it provides recourse for unaware buyers who get swindled.

## 2 A logic of partial awareness

In this section, we introduce our logic of partial awareness.

### 2.1 Syntax

The syntax of our logic has the following components:

- A countable set  $\mathcal{O}$  of constant symbols, representing objects. Following Levesque (1990), we assume that  $\mathcal{O}$  consists of a nonempty set of *standard names*  $d_1, d_2, \dots$ , which may be finite or countably infinite.<sup>4</sup> Intuitively, the standard names will represent the domain elements. We explain the need for these shortly.
- A countably infinite set  $\mathcal{V}^{\mathcal{O}}$  of object variables, which range over objects.
- A countable set  $\mathcal{P}$  of unary predicate symbols.
- A countably infinite set  $\mathcal{V}^{\mathcal{P}}$  of predicate variables.
- A countable set  $\mathcal{C}$  of concept symbols.

If  $a \in \mathcal{O}$ ,  $x \in \mathcal{V}^{\mathcal{O}}$ ,  $P \in \mathcal{P}$ ,  $Y \in \mathcal{V}^{\mathcal{P}}$ , and  $C \in \mathcal{C}$ , then  $P(a)$ ,  $P(x)$ ,  $Y(a)$ ,  $Y(x)$ ,  $C(a)$ , and  $C(x)$  are *atomic formulae*. Starting with these atomic formulae, we construct the set of all formulae recursively: As usual, the set of formulae is closed under conjunction and negation, so if  $\varphi$  and  $\psi$  are formulae, then so are  $\neg\varphi$  and  $\varphi \wedge \psi$ . We allow quantification over objects and over unary predicates, so that if  $\varphi$  is a formula,  $x \in \mathcal{V}^{\mathcal{O}}$ , and  $Y \in \mathcal{V}^{\mathcal{P}}$ , then  $\forall x\varphi$  and  $\forall Y\varphi$  are formulae. Finally, we have two families of modal operators: taking  $\{1 \dots n\}$  to denote the set of agents, we have modal operators  $A_1, \dots, A_n$  and  $K_1, \dots, K_n$ , representing awareness and (explicit) knowledge, respectively. Thus, if  $\varphi$  is a formula, then so is  $A_i\varphi$  and  $K_i\varphi$ . Let  $\mathcal{L}(\mathcal{O}, \mathcal{P}, \mathcal{C})$  denote the the resulting language.

A *simple Boolean formula* is a Boolean combination of formulas of the form  $P(x)$  where  $P \in \mathcal{P}$  and  $x \in \mathcal{V}^{\mathcal{O}}$  that mentions only a single object variable. For example  $P(x) \vee \neg Q(x)$  is a simple Boolean formula, whereas  $P(x) \vee Q(y)$ ,  $P(x) \vee \neg C(x)$ ,  $P(x) \vee \neg K_i P(x)$ , and  $\forall Y(Y(x) \vee \neg Q(y))$  are not. Let  $\mathcal{L}^{sbf}$  denote all such formulae. If  $\varphi(x) \in \mathcal{L}^{sbf}$ , let  $\varphi(a)$  (respectively,  $\varphi(y)$ ) denote the formula where each instance of  $x$  in  $\varphi(x)$  is replaced by  $a \in \mathcal{O}$  (resp.  $y \in \mathcal{V}^{\mathcal{O}}$ ).

<sup>4</sup>Levesque required there to be infinitely many standard names.

For  $\mathcal{Q} \subseteq \mathcal{P}$ , let  $\mathcal{L}^{sbf}(\mathcal{Q})$  be the set of simple Boolean formula that refer only to properties in  $\mathcal{Q}$ .

## 3 Semantics

A model over the language  $\mathcal{L}(\mathcal{O}, \mathcal{P}, \mathcal{C})$  has to give meaning to each of the syntactic elements in the language. We use the standard possible-worlds semantics of knowledge. Thus, a model includes a set  $\Omega$  of possible *states* or *worlds* (we use the two worlds interchangeably) and, for each agent  $i$ , a binary relation  $\mathcal{K}_i$  on worlds. The intuition is that  $(\omega, \omega') \in \mathcal{K}_i$  (sometimes denoted  $\omega' \in \mathcal{K}_i(\omega)$ ) if, in world  $\omega$ , agent  $i$  considers  $\omega'$  possible.

Following HR, we assume that each state  $\omega$  is associated with a language. Formally, there is a function  $\Phi$  on states such that  $\Phi(\omega) = \mathcal{O}_\omega \cup \mathcal{P}_\omega \cup \mathcal{C}_\omega \subseteq \mathcal{O} \cup \mathcal{P} \cup \mathcal{C}$ , where  $\mathcal{O}_\omega = \mathcal{O}$ . We discuss the reason for associating a language with each state below. Let  $\mathcal{L}(\Phi(\omega))$  denote the language associated with state  $\omega$ . We also assume that associated with each state  $\omega$  and agent  $i$ , there is the set of constant, predicate and concept symbols that the agent is aware of; this is given by the function  $\mathcal{A}$ . At state  $\omega$ , each agent can only be aware of symbols that are in  $\Phi(\omega)$ . Thus,  $\mathcal{A}_i(\omega) \subseteq \Phi(\omega)$ . We assume that all agents are aware of the standard names at every state, so that  $\mathcal{A}(\omega)$  includes  $\mathcal{O}$ . We further assume that agents know what they are aware of, so that if  $\omega' \in \mathcal{K}_i(\omega)$ , then  $\mathcal{A}_i(\omega) = \mathcal{A}_i(\omega')$ .

Like Levesque (Levesque 1990), we take the domain  $D$  of a model over  $\mathcal{L}(\mathcal{O}, \mathcal{P}, \mathcal{C})$  to consist of the standard names in  $\mathcal{O}$ . An interpretation  $I$  assigns meaning to the constant and predicate symbols in each state; more precisely, for each state  $\omega$ , we have a function  $I_\omega$  taking  $\mathcal{O}$  to elements of the domain  $D$ ,  $\mathcal{P}$  to subsets of  $D$ , and  $\mathcal{C}$  to simple Boolean formulae. We write  $c_\omega^I$  rather than  $I_\omega(c)$ , etc. To be precise, we have  $c_\omega^I \in D$ ,  $P_\omega^I \subseteq D$ , and  $C_\omega^I \in \mathcal{L}^{sbf}(\mathcal{P}_\omega)$ . We assume that standard names are mapped to themselves, so that  $(d_i)_\omega^I = d_i$ . Notice that the formula defining  $C$  in state  $\omega$  must be expressible in  $\mathcal{L}(\Phi(\omega))$ , the language of  $\omega$ .

Finally, a *valuation*  $V$  is a function that assigns meaning to the object and predicate variables: for each  $x \in \mathcal{V}^{\mathcal{O}}$ , we have  $x^V \in D$ , so that object variables are assigned to elements of the domain (i.e., standard names); and for each  $Y \in \mathcal{V}^{\mathcal{P}}$ ,  $Y^V \in \mathcal{L}^{sbf}(\mathcal{P})$ , so that predicate variables are assigned quantifier-free formulae.

Putting this together, a model for partial awareness has the form

$$M = \langle \Omega, D, \Phi, \mathcal{A}_1 \dots, \mathcal{A}_n, \mathcal{K}_1, \dots, \mathcal{K}_n, I, V \rangle.$$

The truth of a formula  $\varphi \in \mathcal{L}(\mathcal{O}, \mathcal{P}, \mathcal{C})$  at a state  $\omega$  in  $M$ , given a valuation  $V$ , is defined recursively as follows.

- $(M, \omega, V) \models P(a)$  iff  $P(a) \in \mathcal{L}(\Phi(\omega))$  and  $a_\omega^I \in P_\omega^I$ ,
- $(M, \omega, V) \models P(x)$  iff  $P(x) \in \mathcal{L}(\Phi(\omega))$  and  $x_\omega^V \in P_\omega^I$ ,
- $(M, \omega, V) \models \neg\varphi$  iff  $\varphi \in \mathcal{L}(\Phi(\omega))$  and  $(M, \omega, V) \not\models \varphi$ ,
- $(M, \omega, V) \models (\varphi \wedge \psi)$  iff  $(M, \omega, V) \models \varphi$  and  $(M, \omega, V) \models \psi$ ,
- $(M, \omega, V) \models Y(a)$  iff  $(M, \omega, V) \models Y_\omega^V(a)$ ,
- $(M, \omega, V) \models C(a)$  iff  $C(a) \in \mathcal{L}(\Phi(\omega))$  and  $(M, \omega, V) \models C_\omega^I(a)$ ,

- $(M, \omega, V) \models C(x)$  iff  $C(x) \in \mathcal{L}(\Phi(\omega))$  and  $(M, \omega, V) \models C_\omega^I(x)$ ,
- $(M, \omega, V) \models \forall x \varphi$  iff  $(M, \omega, V) \models \varphi[x/d]$  for all constant symbols  $d \in \mathcal{O}$ , where  $\varphi[x/d]$  denotes the result of replacing all free occurrences of  $x$  in  $\varphi$  by  $d$ ,
- $(M, \omega, V) \models \forall Y \varphi$  iff  $(M, \omega, V) \models \varphi[Y/\psi]$ , where  $\psi \in \mathcal{L}^{sbf}(\Phi(\omega))$ ,
- $(M, \omega, V) \models A_i \varphi$  iff  $\varphi \in \mathcal{L}(\mathcal{A}_i(\omega))$ ,
- $(M, \omega, V) \models K_i \varphi$  iff  $(M, \omega, V) \models A_i \varphi$  and  $(M, \omega', V) \models \varphi$  for all  $\omega' \in \mathcal{K}_i(\omega)$ .

Note that what we are calling knowledge here is what has been called *explicit knowledge* in earlier work (Fagin and Halpern 1988; Halpern and R go 2009; 2013). For agent  $i$  to know a formula  $\varphi$ ,  $i$  must also be aware of it. Traditionally,  $K_i$  has been reserved for *implicit* knowledge (where no awareness has been required), and  $X_i$  has been used to denote explicit knowledge (where  $X_i \varphi$  is defined as  $K_i \varphi \wedge A_i \varphi$ ). Since we do not use implicit knowledge in this paper, we have decided to use the more mnemonic  $K_i$  for knowledge, even though it represents explicit knowledge.

For the remainder of the paper, we restrict to models where agents know what they are aware of and knowledge essentially satisfies what are called the S5 properties. Specifically, we restrict to models  $M$  where each  $\mathcal{K}_i$  is an equivalence relation and if  $\omega \in \mathcal{K}_i(\omega')$ , then  $\mathcal{A}_i(\omega) = \mathcal{A}_i(\omega')$ . This implies that  $\mathcal{A}_i(\omega) \subseteq \mathcal{L}(\Phi(\omega'))$  and  $\mathcal{A}_i(\omega') \subseteq \mathcal{L}(\Phi(\omega))$ . Since  $\mathcal{K}_i$  is an equivalence relation, it partitions the states in  $\Omega$ . Thus, we can define  $\mathcal{K}_i$  by describing the partition. (In the economics literature, this partition is called *i's information partition*.)

Given these assumptions, we can now explain why we need different languages at different states. Consider an agent who considers it possible that she is aware of the whole language. Thus, the agent considers possible a state  $\omega'$  such that  $\mathcal{A}_i(\omega') = \mathcal{L}(\Phi(\omega'))$ . If we used the same language at all states, because the agent knows what she is aware, that would mean that, at all states that the agent considered possible, she would know the whole language. Thus, if an agent is aware of all formulas, then she would know that she is aware of all formulas. This is a rather unreasonable property of awareness. It was precisely to avoid this property that Halpern and Rego (2013) allowed different languages to be associated with different states.

We can also explain our use of standard names. Note that to give semantics to  $\forall x \varphi$  and  $\forall Y \varphi$ , we do syntactic replacements. In the case of  $\forall x \varphi$ , we consider all ways of replacing  $x$  by a constant; in the case of  $\forall Y \varphi$ , we consider all ways of replacing  $Y$  by a simple Boolean formula. This is critical because we define awareness syntactically. Consider a formula such as  $\forall x A_i(P(x))$ , and suppose that the agent  $i$  is aware of  $P$ . The standard approach to giving semantics to such a quantified formula would say that  $(M, \omega, V) \models \forall A_i(P(x))$  if  $(M, \omega, V') \models A_i(P(x))$  for all valuations  $V'$  that agree with  $V$  except possibly on  $x$ . But if agent  $i$  is aware of  $P$ , then  $i$  is aware of  $A_i(P(x))$  no matter what  $V'$  is, even if the agent is not aware of some constant  $c$ , and hence not aware of  $A_i(P(c))$ . With our semantics,  $\forall x(P(x))$  is not true un-

less the agent is aware of  $P(c)$  for all constants  $c$ . But if the language had only one constant symbol  $c$ , then  $\forall x A_i(P(x))$  would hold exactly if the agent is aware of  $P(c)$  even if, intuitively, the agent is not aware of all elements of the domain. The use of standard names provides a way of dealing with this problem, by providing us with a way of referring to all domain elements.

Finally, we can explain the role of concepts. Note that the truth of both  $C(a)$  and  $Y(a)$  for some concept  $C$  and predicate variable  $Y$  at a state  $\omega$  are defined by the truth of associated simple Boolean formulae ( $C_\omega^I$  and  $V(Y)$ ), however, as the notation suggests,  $C_\omega^I$  is state-dependent, while  $V(Y)$  is not. Hence, a free predicate variable is semantically a placeholder for a specific fixed formula while a concept can refer to different formulae in different states. Thus, by using concepts, we allow agents to be uncertain about what a label refers to (what exactly are the properties of a quantum computer?) and allow agents to disagree about the relationship between properties and concepts.

A formula  $\varphi$  is *satisfiable* if there exists a model  $M$ , and a state thereof,  $\omega$ , and an assignment  $V$ , such that  $(M, \omega, V) \models \varphi$ . If  $(M, \omega, V) \models \varphi$  for every valuation  $V$ , write  $(M, \omega) \models \varphi$ . Given a model,  $M$ ,  $\varphi$  is *valid* in  $M$ , denoted as  $M \models \varphi$ , if  $(M, \omega) \models \varphi$  for all  $\omega \in \Omega$  such that  $\varphi \in \mathcal{L}(\Phi(\omega))$ . Likewise, for some class of models  $\mathcal{N}$ ,  $\varphi$  is *valid* in  $\mathcal{N}$ , denoted  $\mathcal{N} \models \varphi$ , if  $N \models \varphi$  for all  $N \in \mathcal{N}$ . Note that when we consider the validity of  $\varphi$ , we require only that  $\varphi$  be true in states  $\omega$  such that  $\varphi \in \mathcal{L}(\Phi(\omega))$ . Without this requirement, no formula  $\varphi$  that mentioned any constant, predicate, or concept symbols would be valid, since  $\varphi$  cannot be true at a state  $\omega$  if  $\varphi \notin \mathcal{L}(\Phi(\omega))$ .

This completes our description of the syntax and semantics. Our language is quite expressive. Among other things, we can faithfully embed in it the propositional approach consider by HR and the object-based awareness approach of Board and Chung (2009). In more detail, HR consider a propositional logic of knowledge and awareness, which allows existential quantification over propositions, so has formulas of the form  $\exists X (A_i(X) \wedge \neg A_j(X))$  (there is a formula that agent  $i$  is aware of that agent  $j$  is not aware of). We can capture the HR language by replacing each primitive proposition  $p$  by the atomic formula  $P(d)$ , for some fixed standard name  $d$ , and similarly replacing each proposition variable  $X$  by  $X(d)$ . This replacement allows us to convert a formula  $\varphi$  in the HR language to a formula  $\varphi^r$  in our language (the HR language has no analogue of concepts). We can then convert an HR model  $M$  to a model  $M^r$  in our framework by using the same set of states, taking  $\mathcal{O} = \{d\}$ , and taking  $\mathcal{C} = \emptyset$ , so that there are no concepts and a single object. It is easy to see that a formula  $(M, \omega) \models \varphi$  iff  $(M^r, \omega) \models \varphi^r$ . We can also accommodate the object-based awareness models of Board and Chung (2009). Here the construction is more straightforward: a model where  $\mathcal{C} = \emptyset$  and  $\mathcal{P}_\omega = \mathcal{P}$  for all state  $\omega$  will do the trick.

#### 4 Utility under introspective unawareness

In order to explore how the appeal to concepts might be valuable in a contracting environment, we must add a bit of structure to our problem, dictating the agents' preferences when

they are unaware. This is more complicated than in previous work because we have unawareness of properties. This means that unless the agent knows that she is aware of all properties, it is always possible there exists some property that she is unaware of. As we shall see, this can have enormous consequences for utility. A sufficiently skeptical (and pessimistic) agent might always prefer to avoid objects that she believes satisfy properties that she is unaware of, even if the well-understood consequences are very much in her interest.<sup>5</sup>

We focus on a setting that makes sense for contracting: namely, one where an agent's utility is determined by which set of objects he ends up with. We further simplify things by assuming that the utility of a set of objects is separable, so it is the sum of the utility of the individual objects. Thus, there is no complementarity or substitutability (e.g., it is not the case that the objects are a left shoe and a right shoe, so that having one shoe is useless, while having both has high utility); each agent's preferences can be characterized by a utility function defined on objects.

In this section, we consider various assumptions on the utility function, which, roughly speaking, correspond to different ways of saying that all that an agent cares about are the properties and concepts in the agent's language that the objects satisfy. In the next section, we consider the consequences of these assumptions on a contracting scenario.

Fix a model  $M = \langle \Omega, D, \Phi, \mathcal{A}_1, \dots, \mathcal{A}_n, \mathcal{K}_1, \dots, \mathcal{K}_n, I, V \rangle$  over a language  $\mathcal{L}(\mathcal{O}, \mathcal{P}, \mathcal{C})$ . Let  $\mathcal{U} = \{U_{i,\omega} : D \rightarrow \mathbb{R}\}_{i \in \{1, \dots, n\}, \omega \in \Omega}$  describe the agents' preferences; specifically,  $U_{i,\omega}$  describes agent  $i$ 's preferences in world  $\omega$  by associating with each object its utility (a real number).

Define, for notational expediency, the maps  $\text{PROP}_\omega : D \rightarrow \mathcal{P}$  and  $\text{CON}_\omega : D \rightarrow \mathcal{C}$  by taking  $\text{PROP}_\omega(d) = \{P \in \mathcal{P}_\omega \mid d \in P^I\}$  and  $\text{CON}_\omega(d) = d \mapsto \{C \in \mathcal{C}_\omega \mid d \in C^I\}$ . Thus,  $\text{PROP}_\omega$  and  $\text{CON}_\omega$  take a domain element to the set of properties (resp., concepts) it satisfies at  $\omega$ . Further define  $\text{PROP}_\omega^{A_i}(d) = \text{PROP}_\omega(d) \cap A_i(\omega)$  and  $\text{CON}_\omega^{A_i}(d) = \text{CON}_\omega(d) \cap A_i(\omega)$ ; thus  $\text{PROP}_\omega^{A_i}$  and  $\text{CON}_\omega^{A_i}$  are the restrictions of  $\text{PROP}_\omega$  and  $\text{CON}_\omega$  to agent  $i$ 's awareness.

Our assumptions relate the agents' preferences over domain elements to the properties (and concepts) that these elements possess. Because we allow for introspection, it is possible for an agent to care about aspects of an object that she is unaware of, although she will not be able to articulate exactly why she has such a preference. Thus, we take as a starting point the minimal restriction ensuring an agent's subjective valuation depends only on properties and concepts that the agent can articulate.

- A1. For all  $d, d' \in D$  and all states  $\omega, \omega' \in \Omega$ , if  $\text{PROP}_\omega(d) = \text{PROP}_{\omega'}(d')$  then  $U_{i,\omega}(d) = U_{i,\omega'}(d')$ .

A1 can be viewed as the conjunction of two assumptions: first, if two different objects  $d$  and  $d'$  possess exactly the same properties at a state  $\omega$ , then they are valued identically at  $\omega$ ; second, if a given object  $d$  has the same properties at two different states  $\omega$  and  $\omega'$ , then  $d$  has the same value at

both  $\omega$  and  $\omega'$ . Since each concept is defined (at a possible state) as a simple boolean formula, if two objects satisfy the same properties they must also be instances of the same concepts. Thus, A1 is equivalent to the assumption that if  $\text{PROP}_\omega(d) = \text{PROP}_{\omega'}(d)$  and  $\text{CON}_\omega(d) = \text{CON}_{\omega'}(d')$  then  $U_{i,\omega}(d) = U_{i,\omega'}(d')$ .

Under A1, an agent does not care about the *label* assigned to an object—if  $d$  and  $d'$  satisfy the same properties (and so the same concepts) the agent does not care that  $d$  is called “ $d$ ” and  $d'$  called “ $d'$ ”. A1 also rules out social preferences, or preferences that depend on epistemic conditions. For example, A1 does not allow agent  $i$  to value an object  $d$  according to agent  $j$ 's valuation, or even agent  $j$ 's current knowledge of  $i$ 's valuations, although this might be relevant if  $i$  is interested in reselling  $d$  to  $j$ . Finally, A1 rules out the case where an agent values the same properties differently in different states.

Nonetheless, A1 does allow an agent to value  $d$  and  $d'$  differently even though she can express no distinction between  $d$  and  $d'$  (conditional on the state); the agent might be unaware of all of the ways in which  $d$  and  $d'$  differ. Our next assumption reduces this flexibility in valuations, mandating that an agent's valuation depends only on aspects of the state of which she is aware.

- A2. If  $\text{PROP}_\omega^{A_i}(d) = \text{PROP}_{\omega'}^{A_i}(d')$  and  $\text{CON}_\omega^{A_i}(d) = \text{CON}_{\omega'}^{A_i}(d')$  then  $U_{i,\omega}(d) = U_{i,\omega'}(d')$ .

A2 says that the DM's valuation of objects cannot differ unless the objects are distinguished in some way the agent is aware of. Thus, if two objects are the same in every way that the agent can articulate, then she assigns them the same value. Notice that it is consistent with A2 that an agent  $i$  ascribes different utilities to  $d$  and  $d'$  even though  $i$  is not aware of any property that distinguishes  $d$  and  $d'$ . This can happen if  $d$  and  $d'$  satisfy different concepts. In that case,  $i$  will know that some property must distinguish  $d$  and  $d'$ , but it is a property that  $i$  is not aware of. For instance, an agent might value a quantum computer more than a conventional one even when she does not understand exactly how to define a quantum computer. As this discussion shows, unlike A1, in A2 we must explicitly refer to concepts; even though concepts are built from properties, the agent can be aware of concepts that are defined by properties that the agent is not aware of.

- A3. If  $\text{PROP}_\omega^{A_i}(d) = \text{PROP}_{\omega'}^{A_i}(d')$  then  $U_{i,\omega}(d) = U_{i,\omega'}(d')$ .

A3 says that an agent bases her preferences only on the properties of which she is aware (and not concepts). A3 can be thought of as a principle of neutrality towards unawareness; although two objects are distinguishable (say  $d$  is an instance of  $C$  while  $d'$  is not), the agent values them identically as long as they are not distinguishable by properties that the agent is aware of. That is, while the agent knows there must be a property that separates  $d$  and  $d'$ , because she is unaware of such a property, so she places no value on it. For example, while the agent might understand that there is a difference between classical and quantum computers, because she does not understand *how* these entities differ, she values them the same. It is immediate that A3 implies both

<sup>5</sup>We all probably recall one or two people who, in the early days of mobile technology, refused to carry a cell phone for fear of yet-to-be-discovered health consequences of “radiation”.

A1 and A2. Moreover, in a model without unawareness of properties or concepts, A1, A2, and A3 collapse to the same restriction.

A1 and A2 both seem to us reasonable, while A3 does not seem so reasonable. While it may at first seem unreasonable to base preferences on properties of which you are unaware (as is allowed by A1), it actually is not so uncommon. People may prefer one stock  $d$  to  $d'$  although they cannot articulate a reason that  $d$  is better. Nevertheless, because they see other people buying  $d$  and not  $d'$ , they assume that there is some significant property  $P$  of  $d$  (of which they are not aware) that  $d'$  does not possess that accounts for other people's preferences. Note that this can be modeled using the fact that different states have different languages associated with them. An agent  $i$  might not be aware of  $P$  at a world  $\omega$ , but considers a world  $\omega'$  possible such that  $P(d)$  holds and  $P(d')$  does not. Moreover, he considers  $P$  a “good” property; objects that have property  $P$  get a higher utility than those that do not, all else being equal. We remark that such reasoning certainly seems to play role in the high valuation of some cryptocurrencies! Of course, if  $i$  has a predicate in the language that says “other people like it”, then  $d$  and  $d'$  might be distinguishable using that predicate. This observation emphasizes the fact that A1 and A2 are very much language-dependent. If the agent cannot express various properties in his language, then he may not be able to make distinctions relevant to preferences. (See (Bjorndahl, Halpern, and Pass 2013) for an approach to game theory and decision theory that assumes that utilities are determined by the description of a state in some language.)

## 5 Contracts and conceptual unawareness

We now consider the effect of A1, A2, and A3 on simple interpersonal contracts. Unlike the bulk of the Economics literature, where contracts are functions from a state space to outcomes, we here take a contract to be a *syntactic* object—it makes direct reference to the language of our logic. Real world contracts are syntactic, in that they must literally articulate the contingencies on which they are based. But our motivation for considering syntactic contracts is more than a pursuit of descriptive accuracy. In models of unawareness, the set of contingencies that can be contracted on is a direct consequence of what agents can articulate. So, by considering the language that the agents use, we can directly examine the welfare implications of awareness.<sup>6</sup>

Suppose that we have two agents, 1 and 2. Let  $M$  be the model that describes their uncertainty and awareness, and suppose that their preferences are characterized by  $\mathcal{U}$ . Assume that agent  $i$  is initially endowed with the set of domain

<sup>6</sup>For contracts that do not involve concepts, what we are doing could also be done in a purely semantic framework, for example, that of Heifetz, Meier, and Schipper (2006). However, contracts that involve concepts cannot be expressed in a purely semantic framework, since concepts implicitly involve quantification over syntactic objects (simple Boolean formulas), and thus allow us to get at awareness of unawareness. It seems to us that the use of concepts is indispensable in understanding how unawareness drives novel behavior in contracting environments; that is one of our reasons for modeling unawareness syntactically.

elements  $B_i$ , for  $i = 1, 2$ , where  $B_1$  and  $B_2$  are disjoint. For simplicity, we assume (as is standard) that each agent will consume (i.e., use) exactly one object. Without trade, each agent can consume only an object from her own endowment; with trade, they may be able to do better. Let  $B_1 \oplus B_2$  denote all pairs  $(d_1, d_2)$  of elements in  $B_1 \cup B_2$  such that  $d_1 \neq d_2$ .

A *contract* is a pair  $\langle \Lambda, c \rangle$ , where  $\Lambda$  is a finite set of formulas and  $c$  is a function from  $\Lambda$  to  $B_1 \oplus B_2$ . Let  $c_i$  denote the  $i^{\text{th}}$  projection of  $c$ ; that is, if  $c(\lambda) = (d_1, d_2)$ , then  $c_i(\lambda) = d_i$ . The intuition is that  $c_i$  dictates which object should be consumed by agent  $i$ , contingent on the truth of the formulae in  $\Lambda$ . Given a model  $M$ , and a distinguished state  $\omega^*$ , the contract  $\langle \Lambda, c \rangle$  must satisfy

1.  $M \models \bigvee_{\varphi \in \Lambda} \varphi$ , and
2.  $M \models \neg(\varphi \wedge \psi)$  for all distinct formulae  $\varphi, \psi \in \Lambda$ ,
3.  $\Lambda \subseteq \mathcal{L}(\mathcal{A}_1(\omega^*) \cap \mathcal{A}_2(\omega^*))$ .

The state  $\omega^*$  represents the true state (although the agents may not know which state is actually the true state). The first condition states that some formula in  $\Lambda$  is true in every state (so the contract is complete), the second that the true formula is unique at every state (so the contract is well defined), and the third that both agents are aware of all the statements in the contract (presumably, the act of reading the contract makes them aware all the statements, even if they weren't aware of them beforehand).<sup>7</sup>

It follows from the third condition that  $\Lambda \subseteq \mathcal{L}(\Phi(\omega^*))$ . By conditions 1 and 2, in each state  $\omega$ , there is a unique formula in  $\Lambda$  that is true in  $\omega$ : call that formula  $\varphi_\omega$ . We take the outcome of the contract in state  $\omega$  to be  $c(\varphi_\omega)$ . We abuse notation and write  $c(\omega) = (c_1(\omega), c_2(\omega))$  to denote the outcome of the contract in state  $\omega$ . Thus, the value of the contract to agent  $i$  in state  $\omega$  is  $U_{i,\omega}(c_i(\omega))$ .

We are interested in the value of concepts as a contracting device. To get at this, we must examine the difference between the optimal (or equilibrium) contract when  $\Lambda$  can be any subset of the language and the optimal (or equilibrium) contract when  $\Lambda$  cannot refer to concepts. Given  $M$ ,  $\mathcal{U}$ , and endowments  $B_1, B_2 \subseteq D$ , say that a contract  $\langle \Lambda, c \rangle$  is  $\omega$ -*efficient* if there is no pair of objects  $(d_1, d_2) \in B_1 \oplus B_2$  such that

$$U_{i,\omega}(d_i) \geq U_{i,\omega}(c_i(\omega))$$

for  $i \in \{1, 2\}$ , with at least one inequality strict, and *efficient* if it is  $\omega$ -efficient for all  $\omega \in \Omega$ . In other words, a contract is efficient if it realizes all the gains from trade—if there is no Pareto dominating, feasible assignment of consumption; that is, there is no trade that would leave both agents better off.<sup>8</sup> Finally, call a contract *acceptable for agent  $i$  in state  $\omega$*  if

$$U_{i,\omega'}(c(\omega)) \geq \max_{b \in B_i} U_{i,\omega'}(b_i)$$

<sup>7</sup>The fact agents might not be aware of all statements in a contract before the contract is written clearly has strategic implications. Agent 1 may prefer to leave a clause out of a contract rather than making agent 2 aware of the issue. This issue is studied to some extent by Filiz (2007) and Ozbay (?).

<sup>8</sup>We do not refer to such contracts as Pareto dominant, however, because we are comparing contracts to arbitrary assignments rather than to other contracts.

for all  $\omega' \in \mathcal{K}_i(\omega)$ . Agent  $i$  facing a take-it-or-leave-it offer for an acceptable contract will prefer that contract to her outside option (i.e., consuming an object in  $B_i$ ).

The following examples illustrate how limited awareness can impede the efficiency of contracts and how the reference to concepts can help an agent articulate her preference, tempering the effect of unawareness.

**Example 5.1.** A buyer (agent 1) is trying to purchase a computer from a firm (agent 2).  $B_1 = \{\$$  and  $B_2 = \{c\}$ , where  $\$$  is a fixed amount of money and  $c$  is the computer in question. There are three states:  $\Omega = \{\omega_1, \omega_2, \omega_3\}$ . There are three predicates,  $\mathcal{P} = \{P, Q, R\}$ . In every state,  $R_\omega^I = \{\$$ . We also have  $P_{\omega_1}^I = P_{\omega_2}^I = Q_{\omega_1}^I = \{c\}$  and  $P_{\omega_3}^I = Q_{\omega_2}^I = Q_{\omega_3}^I = \emptyset$ , so that, in the three states,  $c$  has properties  $P$  and  $Q$ , property  $P$ , and no properties, respectively. There is also a single concept, that of a quantum computer,  $QC$ :  $QC_{\omega_1}^I = QC_{\omega_2}^I = P \wedge Q$  and  $QC_{\omega_3}^I = \neg P \wedge \neg Q \wedge \neg R$ . Therefore  $c$  is an instance of  $QC$  in states  $\omega_1$  and  $\omega_3$ . The buyer prefers to purchase the computer if and only if it has property  $Q$ ; the buyer's utility is  $U_{1,\omega_1}(c) > U_{1,\omega_1}(\$)$  and  $U_{1,\omega_k}(\$) > U_{1,\omega_k}(c)$  for  $k \in \{2, 3\}$ . Moreover,  $U_{i,\omega}(\$) = U_{i,\omega'}(\$)$  for all agents  $i$  and states  $\omega, \omega'$ . The firm wants to sell the computer in all states. For now, assume that both agents have full awareness, and their information partitions are given by  $\mathcal{K}_1 = \{\{\omega_1, \omega_2\}, \{\omega_3\}\}$ . Since there is no unawareness, this model trivially satisfies A3 (and hence A1 and A2). Notice that, because the buyer does not know the state, she is unwilling to make any unconditional trade (all constant contracts are unacceptable for the buyer). However, this is easily remedied by the use of a contract. The obvious contract,

$$\begin{aligned} Q(c) &\mapsto c \\ \neg Q(c) &\mapsto \$, \end{aligned}$$

is clearly efficient and acceptable to all parties. ■

Example 1 highlights how contracting can facilitate trade in uncertain environments. Despite the fact that agents do not know which state has obtained, they can eliminate uncertainty by appealing to contracts. The next example illustrates the issues that arise when awareness is limited.

**Example 5.2.** Let  $M$  and  $\mathcal{U}$  be as in Example 1, except that now that agents are not completely aware. Specifically,  $\mathcal{A}_i(\omega) = (\emptyset, \{P, R\}, \emptyset)$  for all  $\omega \in \Omega$  and  $i \in \{1, 2\}$ . Both agents are unaware of  $Q$ . This model satisfies assumptions A1 and A2, but not A3. The contract described in the previous example is no longer articulable. We can circumvent the agents' linguistic limitations by writing a contract in terms of the concept  $QC$ . Indeed, the contract

$$\begin{aligned} P(c) \wedge QC(c) &\mapsto c \\ \neg(P(c) \wedge QC(c)) &\mapsto \$ \end{aligned}$$

implements the same consumption outcomes as the contract in Example 1. ■

In Example 2, the buyer wants to purchase  $c$  only when  $Q(c)$  is true. Since she is unaware of the property  $Q$ , and knows only that there is some property (that is a conjunct

of  $QC$ ) that is desirable, she cannot *directly* demand a computer with property  $Q$ . Before analyzing the contract above, notice if the buyer knew the true state was not  $\omega_3$ , then she could get away with the simple contract that demands  $c$  whenever  $QC(c)$  is true. In states  $\omega_1$  and  $\omega_2$ , the interpretation of  $QC$  is constant, and, given that  $P(c)$  is true in both states,  $QC(c)$  is equivalent to  $Q(c)$  in these states, so the buyer could use the concept of a quantum computer as a proxy for the property  $Q$ .

This simpler contract is not acceptable when the buyer considers all three states possible. In state  $\omega_3$ , the interpretation of a quantum computer is different, so that while  $c$  is an instance of  $QC$  in  $\omega_3$ , it does not satisfy  $Q$ . The buyer is uncertain about the definition of a quantum computer, and while she is unaware of the exact definition in each state, she can articulate the difference: in some states,  $P$  is a property of quantum computers, while in others it is not. By exploiting this difference, she can construct the welfare-optimal contract—she demands  $c$  whenever it possesses the property that defines a quantum computer in addition to  $P$ .

**Example 5.3.** Let  $M$  and  $\mathcal{U}$  be as in Example 2, except for the interpretation of  $QC$  in state  $\omega_1$ . We now set  $QC_{\omega_1}^I = QC_{\omega_3}^I = \neg P \wedge \neg Q \wedge \neg R$ . This model satisfies A1, but not A2 or A3. To see that it violates A2, note that in both  $\omega_1$  and  $\omega_2$ ,  $c$  satisfies  $P$  but not  $R$  nor  $QC$ ; also, in both states  $\$$  satisfies  $R$ , but not  $P$  or  $QC$ . Therefore, if A2 held, the preference ranking between  $c$  and  $\$$  would be the same in both states, which it is not.

It is not hard to see that there is no formula  $\varphi$  such that (a) some agent is aware of  $\varphi$  in both  $\omega_1$  and  $\omega_2$  and (b)  $\varphi$  is true in one of these states and false in the other. We can conclude that there is no efficient contract: any contract that stipulates the sale of  $c$  in state  $\omega_1$  must do the same in state  $\omega_2$ . ■

These examples show how the awareness of agents can affect the set of trading outcomes that are feasible implemented via (syntactic) contracts. Collectively, they suggest a connection between the efficacy of contracting and the relationship between preference and properties as embodied by the various assumptions in the prior section. Under the additional assumption that the agents are aware of the same things in the actual world, this connection can be made formal.

**Theorem 5.1.** *Given a model  $M = \langle \Omega, D, \Phi, \mathcal{A}_1, \dots, \mathcal{A}_n, \mathcal{K}_1, \dots, \mathcal{K}_n, I, V \rangle$ , preferences  $\mathcal{U}$ , endowments  $B_1, B_2 \subseteq D$ , and a state  $\omega^* \in \Omega$  such that  $\mathcal{A}_1(\omega^*) = \mathcal{A}_2(\omega^*)$ , then the following hold:*

- (a) *If  $\langle M, \mathcal{U} \rangle$  satisfies A1 and A2, then there exists a contract  $\langle \Lambda, c \rangle$  that is efficient and acceptable in every state  $\omega' \in \mathcal{K}_1(\omega^*) \cup \mathcal{K}_2(\omega^*)$ .*
- (b) *If  $\langle M, \mathcal{U} \rangle$  satisfies A3, then there exists such a contract with  $\Lambda \subseteq \mathcal{L}^{sbf}$ .*

*Proof.* (Sketch) Define an equivalence relation  $\sim$  on states by taking  $\omega \sim \omega'$  if  $\omega$  and  $\omega'$  agree on all formulas that the agents are aware of in state  $\omega^*$  (recall that the agents are

aware of the same formulae at  $\omega^*$ ); that is,  $\omega \sim \omega'$  if

$$\begin{aligned} \{\varphi \in \mathcal{L} \mid (M, \omega) \models \varphi, (M, \omega^*) \models A_1\varphi\} = \\ \{\varphi \in \mathcal{L} \mid (M, \omega') \models \varphi, (M, \omega^*) \models A_2\varphi\}. \end{aligned}$$

For each  $\omega \in \Omega$ , let  $[\omega]$  denote the  $\omega$ 's cell, that is, the set of states  $\sim$ -equivalent to  $\omega$ . It is straightforward to show that an assignment  $\kappa : \Omega \rightarrow B_1 \oplus B_2$  can be implemented by a contract in state  $\omega^*$  if and only  $\kappa$  works the same way on all states in a cell (since two states in the same cell cannot be distinguished by formulas that the agents are aware of). More precisely, we must have that

$$\kappa^{-1}((d_1, d_2)) \text{ is a union of cells for all } (d_1, d_2) \in B_1 \oplus B_2. \quad (1)$$

Thus, we can view  $\kappa$  as a function from cells to  $B_1 \oplus B_2$ , since it must act the same way on each state in a cell.

For each cell  $L$ , choose an element  $\omega_L \in L$  such that  $\omega_L \in \mathcal{K}_i(\omega^*)$  if  $L \cap \mathcal{K}_i(\omega^*) \neq \emptyset$ ; if  $L \cap \mathcal{K}_i(\omega^*) = \emptyset$ , then just take  $\omega_L = \omega^*$ . Then set

$$\begin{aligned} \kappa(L) \in \max_{(d_1, d_2) \in B_1 \oplus B_2} U_{1, \omega_L}(d_1) \\ \text{such that } U_{2, \omega_L}(d_2) \geq \max_{d \in B_2} U_{2, \omega}(d). \end{aligned}$$

By (1), we can implement this using a contract  $c$ .

Finally, notice that for any two states  $\omega, \omega'$  in the same cell, if  $\mathcal{A}_i(\omega) = \mathcal{A}_i(\omega') = \mathcal{A}_i(\omega^*)$  then  $\text{PROP}_{\omega}^{\mathcal{A}_i} = \text{PROP}_{\omega'}^{\mathcal{A}_i}$  and  $\text{CON}_{\omega}^{\mathcal{A}_i} = \text{CON}_{\omega'}^{\mathcal{A}_i}$ , so by A2,  $U_{i, \omega} = U_{i, \omega'}$ .<sup>9</sup> Hence for all  $\omega \in \mathcal{K}_1(\omega^*) \cup \mathcal{K}_2(\omega^*)$  (recall, agent's know their awareness, so awareness is constant),  $c_\omega$  satisfies the above maximization problem with respect to  $U_{1, \omega}$  and  $U_{2, \omega}$ , so  $c$  is efficient and acceptable.  $\square$

Theorem 5.1(a) says that if agents' preferences can depend only on the properties and concepts that they are aware of, then gains from trade can be fully realized and that the agents will (explicitly) know this. Even if agent might be unaware of some preference-relevant properties, as long as they do not strictly prefer one object to another without being aware of some tangible way in which the objects differ, then they can still articulate an optimal contract. As Example 2 shows, this contract might need to mention concepts. Theorem 5.1(b) says that if agents' preferences depend only on the properties that they are aware of, then they gain nothing from the ability to contract over concepts; there is a contract that they know to be efficient that makes reference only to properties.

If we further relax the restrictions on how the agents assess objects in light of unawareness, contracts that reference concepts may no longer fully compensate for the agents' lack of awareness/articulation. Indeed, as Example 3 shows, we may not be able to write an efficient contract in this case.

**Theorem 5.2.** *There exists a model  $M$ , preferences  $\mathcal{U}$  that satisfy assumption A1, and a state  $\omega$  in  $M$  at which there is no acceptable and efficient contract.*

<sup>9</sup>Note that this is not necessarily true when  $\mathcal{A}_i(\omega) \neq \mathcal{A}_i(\omega')$ , since then  $\omega'$  could model properties that are not in the language of  $\omega$  nor  $\omega^*$ —hence the projection of  $\text{PROP}_{\omega}^{\mathcal{A}_i}$  and  $\text{PROP}_{\omega'}^{\mathcal{A}_i}$  to  $\mathcal{A}_i(\omega^*)$  coincide, but not the functions themselves.

When the agent's have asymmetric unawareness, Theorem 1 does not hold; the gains from trade can be unachievable even when preferences, awareness, and knowledge are all are individually well behaved. The following example illustrates this point.

**Example 5.4.** Let  $B_1 = \{d_1\}$  and  $B_2 = \{d_2\}$ . Consider a language with three predicate symbols  $P$ ,  $Q$ , and  $R$ . Let  $M$  be a model with three states,  $\omega_1$ ,  $\omega_2$ , and  $\omega_3$ , where  $P_{\omega_1}^I = \{d_1\}$ ,  $Q_{\omega_3}^I = \{d_2\}$ ,  $P_{\omega_2}^I = P_{\omega_3}^I = Q_{\omega_1}^I = Q_{\omega_2}^I = \emptyset$ ,  $R_{\omega}^I = \{d_1\}$  for all states  $\omega$ , the only information set for both agents is  $\{\omega_1, \omega_2, \omega_3\}$  (so both agents consider all three worlds possible at all worlds), and  $\mathcal{A}_1(\omega) = (\emptyset, \{P, R\}, \emptyset)$  and  $\mathcal{A}_2(\omega) = (\emptyset, \{Q, R\}, \emptyset)$  for each state  $\omega$ . Now consider what happens when agent 1 wants  $d_1$  only when it has property  $P$  (so wants to trade in states  $\omega_2, \omega_3$ ) and agent 2 wants  $d_2$  only when it has property  $Q$  (so wants to trade in states  $\omega_1, \omega_2$ ). Thus, an efficient contract must induce trade in state  $\omega_2$  and an acceptable contract cannot induce trade except in state  $\omega_2$ . However, neither agent alone can propose such a contract. From agents 1's perspective, states  $\omega_2$  and  $\omega_3$  are indistinguishable in the sense that all objects that he is aware of satisfy the same properties in both states, and from 2's perspective,  $\omega_1$  and  $\omega_2$  are indistinguishable. ■

## 6 A sound and complete axiomatization

We can adapt the axioms used by Halpern and R go (2013) to get a sound and complete axiomatization for our logic. The names of the axioms are, in most cases, the same as or slight modifications of the names used by HR. The axioms are, as usual, relative to a language  $\mathcal{L}(\mathcal{O}, \mathcal{P}, \mathcal{C})$ ; we assume that all the formulas below are in this language.

Consider the following axiom system, which we call AX. Axioms:

Prop. All substitution instances of valid formulas of propositional logic.

AGP.  $A_i\varphi \Leftrightarrow (\bigwedge_{P \in \mathcal{P} \cap \Phi(\varphi)} A_i P(x)) \wedge (\bigwedge_{C \in \mathcal{C} \cap \Phi(\varphi)} A_i C(x))$ , where  $x$  is an arbitrary object variable and  $\Phi(\varphi)$  consists of all the predicate and concept symbols in  $\mathcal{P} \cup \mathcal{C}$  that appear in  $\varphi$ .<sup>10</sup>

KA.  $A_i\varphi \Rightarrow K_i A_i\varphi$

K.  $(K_i\varphi \wedge K_i(\varphi \Rightarrow \psi)) \Rightarrow K_i\psi$ .

T.  $K_i\varphi \Rightarrow \varphi$ .

4.  $K_i\varphi \Rightarrow K_i K_i\varphi$ .

5.  $(\neg K_i\varphi \wedge A_i\varphi) \Rightarrow K_i \neg K_i\varphi$ .

A0.  $K_i\varphi \Rightarrow A_i\varphi$ .

Con.  $\forall X \varphi \Rightarrow \varphi[X/C]$  if  $C$  is a concept.

$1_{\forall_x}$ .  $\forall x \psi \Rightarrow \psi[x/z]$  if  $z$  is either in  $\mathcal{O}$  or is an object variable.

$1_{\forall_X}$ .  $\forall X \varphi \Rightarrow \varphi[X/\psi]$  if  $\psi$  is either in  $\mathcal{L}^{sbf}$  or is a predicate variable.

<sup>10</sup>As usual, the empty conjunction is taken to be vacuously true, so that  $A_i\varphi$  is vacuously true if there are no symbols in  $\mathcal{P} \cup \mathcal{C}$  occur in  $\varphi$ .



$K_{\forall x}. \forall x(\varphi \Rightarrow \psi) \Rightarrow (\forall x\varphi \Rightarrow \forall x\psi).$   
 $K_{\forall X}. \forall X(\varphi \Rightarrow \psi) \Rightarrow (\forall X\varphi \Rightarrow \forall X\psi).$   
 $N_{\forall x}. \varphi \Rightarrow \forall x\varphi$  if  $x$  is not free in  $\varphi$ .  
 $N_{\forall X}. \varphi \Rightarrow \forall X\varphi$  if  $X$  is not free in  $\varphi$ .  
 $\text{Barcan}_x. \forall xK_i\varphi \Rightarrow K_i\forall x\varphi.$   
 $\text{Barcan}_X. (A_i(\varphi) \wedge \forall X(A_i(X(x)) \Rightarrow K_i\varphi) \Rightarrow K_i(\forall X A_i(X(x)) \Rightarrow \forall X\varphi).$   
 $\text{FA}_X. \forall X\neg A_i(X(x)) \Rightarrow K_i(\forall X(\neg A_i(X(x))))).$   
 $\text{Fin}_x. \text{ If } \mathcal{O} = \{c_1, \dots, c_n\}, \text{ then } \forall x\varphi \Leftrightarrow \varphi[x/c_1] \wedge \dots \wedge \varphi[x/c_n].$   
 $\text{Fin}_X. \text{ If } \mathcal{P} \text{ is finite, then } \forall X\varphi \Leftrightarrow \bigwedge_{\psi \in \mathcal{L}^{rep}, \mathcal{P}} \varphi[X/\psi],$   
 where  $\mathcal{L}^{rep}$  is defined and discussed below.

### Rules of Inference:

MP. From  $\varphi$  and  $\varphi \Rightarrow \psi$  infer  $\psi$  (modus ponens).  
 Gen $_K$ . From  $\varphi \wedge A_i\varphi$  infer  $K_i\varphi$ .  
 Gen $_{\forall x}$ . From  $\varphi$  infer  $\forall x\varphi$ .  
 Gen $_{\forall X}$ . If  $P$  is a predicate symbol, then from  $\varphi$  infer  $\forall X\varphi[P/X]$ .

Some comments on the axioms and rules of inference: AGP follows from our assumption that awareness is generated from primitives: agent  $i$  is aware of  $\varphi$  exactly if  $i$  is aware of all the constant, predicate, and concept symbols in  $\varphi$ , and the assumption that agents are aware of all the constant symbols in  $\mathcal{O}$ . KA captures the fact that agents know what they are aware of. T, 4, and 5 are the standard axioms of knowledge (when knowledge is characterized by an equivalence relation), but 5 must be modified to deal with lack of awareness; the approach for doing so goes back to Fagin and Halpern (1988). A0 captures the fact that to (explicitly) know  $\varphi$ ,  $i$  must be aware of  $\varphi$ .  $1_{\forall x}$ ,  $1_{\forall X}$ ,  $N_{\forall x}$ ,  $N_{\forall X}$ ,  $K_{\forall x}$ , and  $K_{\forall X}$  are standard adaptations of axioms for first-order logic, although we remark that the validity of  $1_{\forall x}$  depends on the fact that the constants  $d$  have the same interpretation in all states (otherwise  $\forall xK_i(P(x)) \Rightarrow K_i(P(d))$  would not be valid).

The Barcan $_x$  axiom is a standard axiom of first-order modal logic. The converse is true as well (and provable from the other axioms). However, its validity depends on the fact that all the objects in  $\mathcal{O}$  are in the language of each state. As observed by HR, the analogue for predicate variables does not hold. For example, suppose that the only predicate symbol in  $\Phi(\omega)$  is  $P$  agent  $i$  is aware of  $P$  and  $d$  at  $\omega$ . It is easy to see that (since agents know what they are aware of)  $\forall X K_i A_i(X(d))$  holds at  $\omega$ . However, if  $i$  considers a world  $\omega'$  such that  $\Phi(\omega')$  includes  $Q$  and  $i$  is not aware of  $Q$  at  $\omega'$ , then  $K_i(\forall X A_i(X(d)))$  does not hold at  $\omega$ . As discussed by HR, Barcan $_X$  is essentially the closest approximation to the Barcan formula that we can get in models. To understand it, suppose that  $\varphi$  is a formula with a free predicate variable  $X$ . Then Barcan $_X$  says that if, in a state  $\omega$ , agent  $i$  is aware of all the predicate symbols in  $\varphi$  and if, no matter what formula  $\psi \in \mathcal{L}^{sbf} \cap \Phi(\omega)$  is substituted for  $X$  in  $\varphi$ , if  $i$  is aware of all the predicates in  $\psi$ , then  $i$  knows  $\varphi[X/\psi]$ , then in any state  $\omega'$  that  $i$  considers possible where  $i$  is aware of all predicates

(so that  $\Phi(\omega')$  just includes those predicates that  $i$  is aware of),  $\forall X\varphi$  holds. It is not hard to see that this is sound.

In general, the axiom  $\neg A_i(\varphi) \Rightarrow K_i(\neg A_i\varphi)$  is not valid; If an agent is not aware of a formula, then she can't know it, since we are working with explicit knowledge (even though  $\neg A_i\varphi$  is true in all the worlds that the agent considers possible. FA (which essentially already appears in HR) gives a weak version of that axiom; it says that if the agent is not aware of any predicates, then she knows that. The soundness of Fin $_x$  is clear. Note that if  $\mathcal{O}$  is infinite, there is no corresponding axiom.

For Fin $_X$ , if  $\mathcal{P}$  is finite, define an equivalence relation  $\sim_{\mathcal{P}}$  on formulas in  $\mathcal{L}^{sbf}(\mathcal{P})$  that use the fixed free variable  $x$  by taking  $\varphi \sim_{\mathcal{P}} \psi$  if  $\varphi$  and  $\psi$  are logically equivalent and mention the same predicates. Thus,  $P(x) \wedge Q(x) \sim_{\mathcal{P}} Q(x) \wedge P(x)$ , but  $P(x) \wedge (Q(x) \vee \neg Q(x)) \not\sim_{\mathcal{P}} P(x)$ . There are clearly only finitely many  $\sim_{\mathcal{P}}$  equivalence classes if  $\mathcal{P}$  is finite. Let  $\mathcal{L}^{rep, \mathcal{P}}$  consist of one element from each equivalence class. The soundness of Fin $_X$  can be proved by a straightforward induction on structure. Just as for Fin $_x$ , if  $\mathcal{P}$  is infinite, there is no axiom corresponding to Fin $_X$ .

The rules of inference are all straightforward analogues of the corresponding inference rules in standard first-order logic. However, because of our definition of validity, the soundness of MP is surprisingly hard to prove. Indeed, in HR soundness is proved only in the case that there are infinitely many primitive propositions in the language (the analogue for us would be that  $|\mathcal{P}|$  is infinite); see Lemma A.1, Proposition A.1, and Corollaries A.1 and A.2 in (Halpern and Rêgo 2013). In the supplementary material, we provide a simpler proof of the soundness of MP that works even if the language is finite (and can also be used to prove that MP is sound in the HR setting even if the language is finite).

Modulo these changes, the ideas in the HR proof can be used to prove the soundness and completeness of HR.

**Theorem 6.1.** *AX is a sound and complete axiomatization of  $\mathcal{L}(\mathcal{O}, \mathcal{P}, \mathcal{C})$  with respect to the class of models of partial awareness.*

We leave the details of the proof to the full paper.

## 7 Conclusion

This paper develops and axiomatizes a modal logic, built out of objects, properties, and concepts, to capture unawareness by allowing an agent to be aware of a concept without being aware of the properties that define it. The logic also allows agents to reason about their own unawareness. We show that such a logic is critical for analyzing interpersonal contracts, and that referencing concepts that agents are only partially aware of can improve welfare.

Our analysis of contracts assumed that both agents were aware of all statements in a contract. This makes sense after the contract has been signed, but may well not be true before the contract is written. We believe that an extension of our language to deal with the effects of making other agents aware of certain formulas will allow us explore and analyze the dynamic process of contract writing. We hope to do this in future work.

## References

- Bjorndahl, A.; Halpern, J. Y.; and Pass, R. 2013. Language-based games. In *Theoretical Aspects of Rationality and Knowledge: Proc. Fourteenth Conference (TARK 2013)*, 39–48.
- Board, O., and Chung, K.-S. 2009. Object-based unawareness: theory and applications. Working paper 378, University of Pittsburgh.
- Fagin, R., and Halpern, J. Y. 1988. Belief, awareness, and limited reasoning. *Artificial Intelligence* 34:39–76.
- Filiz, E. 2007. Incorporating unawareness into contract theory. In *Theoretical Aspects of Rationality and Knowledge: Proc. Eleventh Conference (TARK 2007)*, 135–144.
- Halpern, J. Y., and Rêgo, L. C. 2009. Reasoning about knowledge of unawareness. *Games and Economic Behavior* 67(2):503–525.
- Halpern, J. Y., and Rêgo, L. C. 2013. Reasoning about knowledge of unawareness revisited. *Mathematical Social Sciences* 66(2):73–84.
- Heifetz, A.; Meier, M.; and Schipper, B. 2006. Interactive unawareness. *Journal of Economic Theory* 130:78–94.
- Levesque, H. J. 1990. All I know: a study in autoepistemic logic. *Artificial Intelligence* 42(3):263–309.
- Modica, S., and Rustichini, A. 1994. Awareness and partitioned information structures. *Theory and Decision* 37:107–124.
- Modica, S., and Rustichini, A. 1999. Unawareness and partitioned information structures. *Games and Economic Behavior* 27(2):265–298.
- Sillari, G. 2008. Quantified logic of awareness and impossible possible worlds. *Review of Symbolic Logic* 1(4):514–529.