Partial Awareness

Joseph Y. Halpern
Computer Science Department
Cornell University
halpern@cs.cornell.edu

Evan Piermont
Economics Department
Royal Holloway, University of London
evan.piermont@rhul.ac.uk

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. We take an agent to be partially aware of a concept if she is aware of the concept without being aware of the properties that define it. The logic allows for quantification over objects and properties, so that the agent can reason about her own unawareness. We then apply the logic to contracts, which we view as syntactic objects that dictate outcomes based on the truth of formulas. We show that when agents are unaware of some relevant properties, 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 φ if i is unaware of φ. 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 φ with an agent neither knowing φ nor knowing that she does not know φ (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 φ with the sentiment "φ is not on my radar screen". With this interpretation, partial awareness becomes "some aspects of φ are on my radar screen".[1] 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 an extreme case. A slightly more aware agent might be aware of (the concept of) quantum computers, having read a magazine article about them. 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 then 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;[2] 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 possesses 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 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 con-

[1] Lack of awareness of φ has also been identified with the inability of compute whether φ 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 φ are true."

[2] 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.

sider 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.

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, for an arbitrary property R, the statement $K_i\big(\forall P \forall x((Q(x) \Leftrightarrow C(x) \wedge P(x)) \Rightarrow (P(x) \Rightarrow R(x))\big)$, 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 has the ability to teach the buyer; specifically, 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 reveal his information in a credible way, 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 in state 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.[3]

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 an 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 $\forall x\big(T(x) \Leftrightarrow C(x) \wedge QP'(x)\big)$, 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 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.

[3]This argument does not rely on the fact that there are only two states. Suppose that there are n states, say $s_1, \ldots, s_n$, and the buyer is willing to pay p_i for the computer if the true state is s_i, with $p_1 \geq p_2 \geq \ldots \geq p_n$. 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.)

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 building blocks:

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

If $d \in \mathscr{O}$, $x \in \mathcal{V}^{\mathscr{O}}$, $P \in \mathscr{P}$, $Y \in \mathcal{V}^{\mathscr{P}}$, and $C \in \mathscr{C}$, then $P(d)$, $P(x)$, $Y(d)$, $Y(x)$, $C(d)$, and $C(x)$ are *atomic formulas*. Starting with these atomic formulas, we construct the set of all formulas recursively: As usual, the set of formulas is closed under conjunction and negation, so if φ and ψ are formulas, then so are $\neg\varphi$ and $\varphi \wedge \psi$. We allow quantification over objects and over unary predicates, so that if φ is a formula, $x \in \mathcal{V}^{\mathscr{O}}$, and $Y \in \mathcal{V}^{\mathscr{P}}$, then $\forall x\varphi$ and $\forall Y\varphi$ are formulas. Finally, we have two families of modal operators: taking $\{1 \ldots n\}$ to denote the set of agents, we have modal operators $A_1, \ldots, A_n$ and $K_1, \ldots, K_n$, representing awareness and (explicit) knowledge, respectively. Thus, if φ is a formula, then so is $A_i\varphi$ and $K_i\varphi$. Let $\mathcal{L}(\mathscr{O}, \mathscr{P}, \mathscr{C})$ denote the resulting language. A formula that contains no free variables is called a *sentence*.

2.2 Semantics

A model over the language $\mathcal{L}(\mathscr{O}, \mathscr{P}, \mathscr{C})$ 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 Ω of possible *states* or *worlds* (we use the two words 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 ω, agent i considers ω' possible.

[4] Levesque required there to be infinitely many standard names.

Following HR, we assume that each state ω is associated with a language. Formally, there is a function Φ on states such that $\Phi(\omega) = (\mathscr{O}_\omega, \mathscr{P}_\omega, \mathscr{C}_\omega)$, where $\mathscr{O}_\omega = \mathscr{O}$, $\mathscr{P}_\omega \subseteq \mathscr{P}$, and $\mathscr{C}_\omega \subseteq \mathscr{C}$. We discuss the reason for associating a language with each state below. Let $\mathcal{L}(\Phi(\omega))$ denote the language associated with state ω. We also assume that associated with each state ω 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 ω, 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 $\mathscr{O}$.

Like Levesque (1990), we take the domain D of a model over $\mathcal{L}(\mathscr{O}, \mathscr{P}, \mathscr{C})$ to consist of the standard names in $\mathscr{O}$. An interpretation I assigns meaning to the constant and predicate symbols in each state; more precisely, for each state ω, we have a function I_ω taking $\mathscr{O}$ to elements of the domain D, $\mathscr{P}$ to subsets of D, and $\mathscr{C}$ to Boolean combinations of properties (i.e., predicates). This last item requires some explanation. Although elements of $\mathscr{O}$ and $\mathscr{P}$ are mapped to semantic objects (elements in the domain and sets of elements in the domain, respectively), elements of $\mathscr{C}$ are mapped to *syntactic* objects: Boolean combinations of properties. Let $\mathcal{L}^{bc}$ denote the Boolean combination of properties; if $\mathscr{P}' \subseteq \mathscr{P}$, let $\mathcal{L}^{bc}(\mathscr{P}')$ denote the Boolean combination of properties in $\mathscr{P}'$. We require that $\mathcal{I}_\omega(C) \in \mathcal{L}^{bc}(\Phi(\omega))$, so that the Boolean combination defining C in state ω must be expressible in $\mathcal{L}(\Phi(\omega))$, the language of ω. We sometimes write c^I_ω rather than $I_\omega(c)$, P^I_ω rather than $I_\omega(P)$, and C^I_ω rather than $I_\omega(C)$. We assume that standard names are mapped to themselves, so that $(d_i)^I_w = d_i$.

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

$$M = (\Omega, D, \Phi, \mathcal{A}_1 \ldots, \mathcal{A}_n, \mathcal{K}_1, \ldots, \mathcal{K}_n, I).$$

The truth of a sentence $\varphi \in \mathcal{L}(\mathscr{O}, \mathscr{P}, \mathscr{C})$ at a state ω in M is defined recursively as follows.

- $(M, \omega) \models P(d)$ iff $P(d) \in \mathcal{L}(\Phi(\omega))$ and $d \in P^I_\omega$,
- $(M, \omega) \models \neg\varphi$ iff $\varphi \in \mathcal{L}(\Phi(\omega))$ and $(M, \omega) \not\models \varphi$,
- $(M, \omega) \models (\varphi \wedge \psi)$ iff $(M, \omega) \models \varphi$ and $(M, \omega) \models \psi$,
- $(M, \omega) \models C(d)$ iff $C(d) \in \mathcal{L}(\Phi(\omega))$ and $(M, \omega) \models C^I_\omega(d)$,
- $(M, \omega) \models \forall x\varphi$ iff $(M, \omega) \models \varphi[x/d]$ for all constant symbols $d \in \mathscr{O}$, where $\varphi[x/d]$ denotes the result of replacing all free occurrences of x in φ by d,
- $(M, \omega) \models \forall Y\varphi$ iff $(M, \omega) \models \varphi[Y/\psi]$, where $\psi \in \mathcal{L}^{bc}(\Phi(\omega))$,[5]
- $(M, \omega) \models A_i\varphi$ iff $\varphi \in \mathcal{L}(\mathcal{A}_i(\omega))$,
- $(M, \omega) \models K_i\varphi$ iff $(M, \omega) \models A_i\varphi$ and $(M, \omega') \models \varphi$ for all $\omega' \in \mathcal{K}_i(\omega)$.

[5] There is an abuse of notation here. For example, if ψ is $P \wedge (Q \vee R)$ and φ is $Y(d)$, then $\varphi[Y/\psi]$ is $P(d) \wedge (Q(d) \vee R(d))$; that is, we apply the arguments of φ to all predicates in the ψ. We hope that the intended formula is clear in all cases.

Note that what we are calling knowledge here is what has been called *explicit knowledge* in earlier work (Fagin and Halpern 1988; Halpern and Rêgo 2009; 2013): for agent i to know a formula φ, 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 Ω. Thus, we can define $\mathcal{K}_i$ by describing the partition. (In the economics literature, this partition is called i's *information partition*.) For the remainder of the paper, we assume (as is standard in the literature) that agents know what they are aware of, so that if $\omega \in \mathcal{K}_i(\omega')$, then $\mathcal{A}_i(\omega) = \mathcal{A}_i(\omega')$, and that each $\mathcal{K}_i$ is an equivalence relation, and thus partitions the states in Ω. We can define $\mathcal{K}_i$ by describing this partition, called i's *information partition* in the economics literature.

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 ω' such that $\mathcal{A}_i(\omega') = \Phi(\omega')$. If we used the same language at all states, because the agent knows what she is aware of, 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 Boolean combination of properties. This is critical because we define awareness syntactically. Consider a formula such as $\forall Y A_i(Y(d))$. The standard approach to giving semantics to such a quantified formula would say, roughly speaking, that $(M,\omega) \models \forall Y A_i(Y(d))$ if $(M,\omega) \models A_i(Y(d))$ no matter which set of objects Y represents. The "no matter which property (i.e., set of objects) Y represents" would typically be captured by including a valuation V on the left-hand side of $\models$, where V interprets Y as a set of objects; we would then consider all valuations V that agree on their interpretations of all predicate variables but Y. Since we treat awareness syntactically, this approach will not work. We need to replace Y by a syntactic object and then evaluate whether agent i is aware of the resulting formula. This is exactly what we do: $(M,\omega) \models A_i(Y(d))$ if $(M,\omega) \models A_i(\psi(d))$ for $\psi \in \mathcal{L}^{bc}(\Phi(\omega))$.

A sentence φ is *satisfiable* if there exists a model M and a state ω in M such that $(M,\omega) \models \varphi$. Given a model M, φ is *valid* in M, denoted $M \models \varphi$, if $(M,\omega) \models \varphi$ for all $\omega \in \Omega$ such that $\varphi \in \mathcal{L}(\Phi(\omega))$. Likewise, for some class of models $\mathcal{N}$, φ 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 φ, we follow Halpern and Rêgo (2013) in requiring only that φ be true in states ω such that $\varphi \in \mathcal{L}(\Phi(\omega))$. Thus, φ is valid if φ is true in all states ω where φ is part of the language of ω; we are not interested in whether φ is true if φ is not in the language (indeed, φ is guaranteed not to be true in this case).

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 considered 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 replacing each proposition variable X by $X(d)$. This replacement allows us to convert a formula φ in the HR language to a sentence φ^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 $\mathscr{O} = \{d\}$, and taking $\mathscr{C} = \emptyset$, so that there are no concepts and a single object. It is easy to see that $(M,\omega) \models \varphi$ 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 $\mathscr{C} = \emptyset$ and $\mathscr{P}_\omega = \mathscr{P}$ for all states ω will do the trick.

3 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. Thus, unless the agent knows that she is aware of all properties, it is always possible there exists some property that she is unaware of.

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

Define, for notational expediency, the maps $\textsc{prop}_\omega : D \to \mathscr{P}$ and $\textsc{con}_\omega : D \to \mathscr{C}$ by taking $\textsc{prop}_\omega(d) = \{P \in \mathscr{P}_\omega : d \in P^I_\omega\}$ and $\textsc{con}_\omega(d) = \{C \in \mathscr{C}_\omega : d \in C^I_\omega\}$. Thus, $\textsc{prop}_\omega$ and $\textsc{con}_\omega$ take a domain element to the set of properties (resp., concepts) it satisfies at ω. Further define $\textsc{prop}^{\mathcal{A}_i}_\omega(d) = \textsc{prop}_\omega(d) \cap \mathcal{A}_i(\omega)$ and $\textsc{con}^{\mathcal{A}_i}_\omega(d) = \textsc{con}_w(d) \cap \mathcal{A}_i(\omega)$; thus $\textsc{prop}^{\mathcal{A}_i}_\omega$ and $\textsc{con}^{\mathcal{A}_i}_\omega$ are the restrictions of $\textsc{prop}_w$ and $\textsc{con}_w$ 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 $\textsc{prop}_\omega(d) = \textsc{prop}_{\omega'}(d')$ 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 ω, then they are valued identically at ω; second, if a given object d has the same properties at two different states ω and ω', then d has the same value at both ω and ω'. Since each concept is defined (at a possible state) as a Boolean combination of properties, if two objects satisfy the same properties then they must also be instances of the same concepts. Thus, A1 is equivalent to the assumption that if $\textsc{prop}_\omega(d) = \textsc{prop}_{\omega'}(d)$ and $\textsc{con}_\omega(d) = \textsc{con}_{\omega'}(d')$ 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.

A1 allows an agent to value d and d' differently even if she can express no distinction between d and d' (conditional on the state. This can happen if d and d' differ on properties that the agent is unaware of. 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 $\textsc{prop}^{\mathcal{A}_i}_\omega(d) = \textsc{prop}^{\mathcal{A}_i}_{\omega'}(d')$ and $\textsc{con}^{\mathcal{A}_i}_\omega(d) = \textsc{con}^{\mathcal{A}_i}_{\omega'}(d')$ 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. If two objects are the same in every way that the agent can articulate, then she assigns them the same value. 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 knows 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 $\textsc{prop}^{\mathcal{A}_i}_\omega(d) = \textsc{prop}^{\mathcal{A}_i}_{\omega'}(d')$ 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 equally. It is immediate that A3 implies both A1 and A2. Moreover, in a model without unawareness of properties or concepts, A1, A2, and A3 collapse to the same restriction.

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 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. 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 ω, but considers a world ω' 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 a 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.)

4 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.[6]

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 elements End_i, for $i = 1, 2$, where End_1 and End_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 $End_1 \oplus End_2$ denote all pairs (d_1, d_2) of elements in $End_1 \cup End_2$ such that $d_1 \neq d_2$. Note that we might have $(d_1, d_2) \in End_1 \cup End_2$ even if d_1 and d_2 are both in End_2 (and not in End_1); the agents may both consume something that was in agent 2's initial endowment.

A *contract* is a pair $\langle \Lambda, c \rangle$, where Λ is a finite set of sentences and c is a function from Λ to $End_1 \oplus End_2$. Let c_i denote the i^{th} component 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 sentences in Λ. Given a model M, 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 sentences $\varphi, \psi \in \Lambda$.

The first condition states that some sentence in Λ is true in every state (so the contract is complete), the second that the true sentence is unique at every state (so the contract is well defined). A contract is *articulable* in state ω^*, sometimes denoted ω^**-articulable*, if $\Lambda \subseteq \mathcal{L}(\mathcal{A}_1(\omega^*) \cap \mathcal{A}_2(\omega^*))$, that is, if 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.[7]) Note that if a contract is articulable in ω^* then $\Lambda \subseteq \mathcal{L}(\Phi(\omega^*))$.

By conditions 1 and 2, in each state ω, there is a unique sentence in Λ that is true in ω: call that sentence φ_ω. We take the outcome of the contract in state ω 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 ω. Thus, the value of the contract to agent i in state ω 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 Λ can be any subset of the language and the optimal (or equilibrium) contract when Λ cannot refer to concepts. Given M, $\mathcal{U}$, and endowments $End_1, End_2 \subseteq D$, say that a contract $\langle \Lambda, c \rangle$ is ω*-efficient* if there is no pair of objects $(d_1, d_2) \in End_1 \oplus End_2$ 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 ω-efficient for all $\omega \in \Omega$. In other words, a contract is efficient if it realizes all the gains from trade, so there is no trade that would leave both agents better off. Finally, a contract is ω*-acceptable for agent* i if

$$U_{i,\omega'}(c(\omega)) \geq \max_{d \in End_i} U_{i,\omega'}(d)$$

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 End_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 4.1. A buyer (agent 1) is trying to purchase a computer from a firm (agent 2). $End_1 = \{d^{\$}\}$ and $End_2 = \{d^{cmp}\}$, where $d^{\$}$ is a fixed amount of money and d^{cmp} is the computer in question. There are three states: $\Omega = \{\omega_1, \omega_2, \omega_3\}$. There are three predicates, $\mathscr{P} = \{P, Q, R\}$. We have $R^I_{\omega_1} = R^I_{\omega_2} = R^I_{\omega_3} = \{d^{\$}\}$; in addition, $P^I_{\omega_1} = P^I_{\omega_2} = Q^I_{\omega_1} = \{d^{cmp}\}$ and $P^I_{\omega_3} = Q^I_{\omega_2} = Q^I_{\omega_3} = \emptyset$, so that, in the three states, d^{cmp} has properties P and Q, property P, and no properties, respectively. There is also a single concept, that of a quantum computer, QC: $QC^I_{\omega_1} = QC^I_{\omega_2} = P \wedge Q$ and $QC^I_{\omega_3} = \neg P \wedge \neg Q \wedge \neg R$. Therefore c is an instance of QC in states ω_1 and ω_3. The buyer prefers to purchase the computer if and only if it has property Q; thus, the buyer's utility is such that $U_{1,\omega_1}(d^{cmp}) > U_{1,\omega_1}(d^{\$})$ and $U_{1,\omega_k}(d^{\$}) > U_{1,\omega_k}(d^{cmp})$ for $k \in \{2, 3\}$. Moreover, $U_{i,\omega}(d^{\$}) = U_{i,\omega'}(d^{\$})$ for all agents i and states ω, ω'. The

[6] 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 represent syntactic objects. Agents can be aware of a concept without being aware of the properties that the concept represents (as in the case of quantum computers); because of this, concepts allow us to indirectly get at awareness of unawareness. Awareness of unawareness seems difficult to express in a purely semantic framework (and, indeed, cannot be expressed in the framework of Heifetz, Meier, and Schipper). It seems to us that the use of concepts is indispensable in understanding how unawareness drives novel behavior in contracting environments; this is one of our reasons for modeling unawareness syntactically. Syntactic contracts were considered by Piermont (2017), with similar motivation.

[7] The fact that 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 (2012) and Ozbay (2007).

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 $\mathcal{K}_1 = \{\{\omega_1, \omega_2, \omega_3\}\}$ and $\mathcal{K}_2 = \{\{\omega_1, \omega_2\}, \{\omega_3\}\}$. Since there is no unawareness, this model trivially satisfies A3 (and hence A1 and A2). 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, where $\Lambda = \{Q(d^{cmp}), \neg Q(d^{cmp})$ and the function c is given by

$$Q(d^{cmp}) \mapsto (d^{cmp}, d^{\$})$$
$$\neg Q(d^{cmp}) \mapsto (d^{\$}, d^{cmp}),$$

is clearly efficient and acceptable to all parties. ∎

Example 4.1 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 4.2. Let M and $\mathcal{U}$ be as in Example 4.1, except that now that agents are not completely aware. Specifically, $\mathcal{A}_i(\omega) = (\mathscr{O}, \{P, R\}, \mathscr{C})$ 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 consider (Λ, c), where $\Lambda = \{P(d^{cmp}) \wedge Q(d^{cmp}), \neg(P(d^{cmp} \wedge Q(d^{cmp}))$ and c is given by

$$P(d^{cmp}) \wedge QC(d^{cmp}) \mapsto (d^{cmp}, d^{\$})$$
$$\neg\big(P(d^{cmp}) \wedge QC(d^{cmp})\big) \mapsto (d^{\$}, d^{cmp}),$$

implements the same consumption outcomes as the contract in Example 4.1. ∎

In Example 4.2, the buyer wants to purchase d^{cmp} only when $Q(d^{cmp})$ 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 ω_3, then she could get away with the simple contract that demands d^{cmp} whenever $QC(d^{cmp})$ is true. In states ω_1 and ω_2, the interpretation of QC is constant, and, given that $P(d^{cmp})$ is true in both states, $QC(d^{cmp})$ is equivalent to $Q(d^{cmp})$ 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 ω_3, the interpretation of a quantum computer is different, so that while d^{cmp} is an instance of QC in ω_3, it does not satisfy Q. The buyer is uncertain about the definition of a quantum computer; 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 d^{cmp} whenever it possesses the property that defines a quantum computer in addition to P.

These examples show how the awareness of agents can affect the set of trading outcomes that can be implemented via (syntactic) contracts. Collectively, they suggest a connection between the efficacy of contracting and the relationship between preference and properties as embodied by assumptions A1–A3. Under the additional assumption that the agents are aware of the same things in the actual world (a reasonable assumption if we assume that the language talks only about contract-relevant features, and both agents have read the contract, so are aware of all the properties and concepts that the contract mentions), this connection is made formal in the following result, whose proof (like that of all other theorems) is left to the full paper, which can be found on arxiv.

Theorem 4.1. *Given a model $M = (\Omega, D, \Phi, \mathcal{A}_1 \ldots, \mathcal{A}_n, \mathcal{K}_1, \ldots, \mathcal{K}_n, I)$, preferences $\mathcal{U}$, endowments $End_1, End_2 \subseteq D$, and a state $\omega^* \in \Omega$ such that $\mathcal{A}_1(\omega^*) = \mathcal{A}_2(\omega^*)$ and $\mathcal{K}_1(\omega^*) \cup \mathcal{K}_2(\omega^*)$ is finite, the following hold:*

(a) If $\langle M, \mathcal{U}\rangle$ satisfies A1 and A2, then there exists a contract that is ω^*-articulable, ω^*-efficient, and ω^*-acceptable for $i = 1, 2$.

(b) If in addition, for $i = 1, 2$, $\mathcal{A}_i$ is constant on $\mathcal{K}_1(\omega^*) \cup \mathcal{K}_2(\omega^*)$, then there exists a contract that is ω^*-articulable, ω'-efficient, and ω'-acceptable for all $\omega' \in \mathcal{K}_1(\omega^*) \cup \mathcal{K}_2(\omega^*)$.

(c) If in addition $\langle M, \mathcal{U}\rangle$ satisfies A3, then there exists a contract $\langle\Lambda, c\rangle$ that is ω^*-articulable, ω'-efficient, and ω'-acceptable for i at all $\omega' \in \mathcal{K}_1(\omega^*) \cup \mathcal{K}_2(\omega^*)$ such that $\Lambda \subseteq \mathcal{L}^{bc}$.

Theorem 4.1(a) says that if agents' preferences can depend only on properties and concepts that they are aware of, then gains from trade can be fully realized. Even if agents are 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 that the objects differ, then they can still articulate an optimal contract. As Example 4.2 shows, this contract might need to mention concepts. Part (b) states that if, in addition, each agent knows what the other is aware of, then each of them knows that gains from trade can be achieved. That is, both agents know that, no matter what the true state of the world is (from their perspective), trading is worthwhile. Theorem 4.1(c) 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.

Theorem 4.1 requires agents to be aware of the same properties in ω^*. As we argued above, this is a reasonable assumption. As seen by the following example, the assumption is also necessary.

Example 4.3. Let $End_1 = \{d_1\}$ and $End_2 = \{d_2\}$. Consider a language with three predicate symbols P, Q, and R. Let M be a model with three states, ω_1, ω_2, and ω_3, where $P^I_{\omega_1} = \{d_1\}$, $Q^I_{\omega_3} = \{d_2\}$, $P^I_{\omega_2} = P^I_{\omega_3} = Q^I_{\omega_1} = Q^I_{\omega_2} = \emptyset$,

$R^I_\omega = \{d_1\}$ for all states ω, 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), $\mathcal{A}_1(\omega) = (\mathscr{O}, \{P, R\}, \emptyset)$, and $\mathcal{A}_2(\omega) = (\mathscr{O}, \{Q, R\}, \emptyset)$ for each state ω. Now consider what happens when agent 1 wants d_1 only when it has property P (so wants to trade in states ω_2, ω_3) and agent 2 wants d_2 only when it has property Q (so wants to trade in states ω_1 and ω_2). Thus, an efficient contract must induce trade in state ω_2 and an acceptable contract cannot induce trade except in state ω_2. However, neither agent alone can propose such a contract. From agents 1's perspective, states ω_2 and ω_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, ω_1 and ω_2 are indistinguishable. ∎

5 Axiomatization and complexity

We can adapt the axioms used by Halpern and Rêgo (2013) to get a sound and complete axiomatization for our logic, provided that the set $\mathscr{P}$ of predicates is infinite. This assumption seems reasonable, given that we are mainly interested in agents who are never sure that they are aware of all predicates. (HR make an analogous assumption.)

Consider the following axiom system, which we call AX.

Axioms:

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

AGP. $A_i\varphi \Leftrightarrow (\wedge_{P\in\mathscr{P}\cap\Phi(\varphi)} A_iP(x)) \wedge (\wedge_{C\in\mathscr{C}\cap\Phi(\varphi)} A_iC(x))$, where x is an arbitrary object variable and $\Phi(\varphi)$ consists of all the predicate and concept symbols in $\mathscr{P} \cup \mathscr{C}$ that appear in φ.[8]

KA. $A_i\varphi \Rightarrow K_iA_i\varphi$

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

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

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

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. $\exists X(\forall x(C(x) \Leftrightarrow X(x)))$.

$1_{\forall_x}$ $\forall x\psi \Rightarrow \psi[x/c]$ for $c \in \mathscr{O}$.

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

$\mathrm{K}_{\forall x}$. $\forall x(\varphi \Rightarrow \psi) \Rightarrow (\forall x\varphi \Rightarrow \forall x\psi)$.

$\mathrm{K}_{\forall X}$. $\forall X(\varphi \Rightarrow \psi) \Rightarrow (\forall X\varphi \Rightarrow \forall X\psi)$.

$\mathrm{N}_{\forall x}$. $\varphi \Rightarrow \forall x\varphi$ if x is not free in φ.

$\mathrm{N}_{\forall X}$. $\varphi \Rightarrow \forall X\varphi$ if X is not free in φ.

Barcan_x. $\forall xK_i\varphi \Rightarrow K_i\forall x\varphi$.

Barcan_X. $(A_i(\forall X\varphi) \wedge \forall X(A_i(X(c))) \Rightarrow K_i\varphi) \Rightarrow K_i(\forall XA_i(X(c)) \Rightarrow \forall X\varphi)$.

FA_X. $\forall X\neg A_i(X(c)) \Rightarrow K_i(\forall X\neg A_i(X(c)))$.

Fin_x. If $\mathscr{O} = \{c_1, \ldots, c_n\}$, then $\forall x\varphi \Leftrightarrow \varphi[x/c_1] \wedge \ldots \wedge \varphi[x/c_n]$.

Rules of Inference:

MP. From φ and $\varphi \Rightarrow \psi$ infer ψ (modus ponens).

Gen_K. From $\varphi \wedge A_i\varphi$ infer $K_i\varphi$.

$\mathrm{Gen}_{\forall_x}$. From φ infer $\forall x\varphi[c/x]$, where $c \in \mathscr{O}$.

$\mathrm{Gen}_{\forall_X}$. If P is a predicate symbol, then from φ infer $\forall X\varphi[P/X]$.

Theorem 5.1. *AX is a sound and complete axiomatization of $\mathcal{L}(\mathscr{O}, \mathscr{P}, \mathscr{C})$ with respect to the class of models of partial awareness, if $\mathscr{P}$ is infinite.*

Since the logic is axiomatizable, the validity problem is recursively enumerable. This is also a lower bound on its complexity, even if we do not allow quantification over predicates, since first-order epistemic logic with just two unary predicates was shown by Kripke (1962) to be undecidable. Kripke's proof used the well-known fact that first-order logic with a single binary predicate R is undecidable, and the observation that $R(x, y)$ can be represented as $\neg K\neg(P(x) \wedge Q(y))$. (We must add the formula $\forall x(A(P(x) \wedge Q(x))$ to ensure that awareness does not cause a problem.) We thus get Thus,

Theorem 5.2. *The validity problem for the language $\mathcal{L}(\mathscr{O}, \mathscr{P}, \mathscr{C})$ in the class of models of partial awareness is r.e.-complete if $|\mathscr{P}| \geq 2$.*

6 Conclusion

We have defined and axiomatized a modal logic that captures partial 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. We believe that the logic should also be applicable to other domains, such an analyzing communication between people. We hope to consider such applications in the future.

We are also interested in analyzing dynamic aspects of awareness, and applying this to contracts. 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.

Acknowledgments: Halpern was supported in part by NSF grants IIS-1703846 and IIS-1718108, ARO grant W911NF-17-1-0592, and a grant from the Open Philanthropy project.

References

Bjorndahl, A.; Halpern, J. Y.; and Pass, R. 2013. Language-based games. In *Theoretical Aspects of Rationality and*

[8]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 $\mathscr{P} \cup \mathscr{C}$ occur in φ.

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-Ozbay, E. 2012. Incorporating unawareness into contract theory. *Games and Economic Behavior* 76:181–194.

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.

Kripke, S. 1962. The undecidability of monadic modal quantification theory. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* 8:113–116.

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 partitional information structures. *Theory and Decision* 37:107–124.

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

Ozbay, E. 2007. Unawareness and strategic announcements in games with uncertainty. 231–238.

Piermont, E. 2017. Introspective unawareness and observable choice. *Games and Economic Behavior* 106:134–152.

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

7 More on contracts

In this section, we expand on the discussion of contracts in Section 4. We start with the proof of Theorem 4.1. For the reader's convenience, we repeat the statement of the theorem.

Theorem 4.1: *Given a model $M = (\Omega, D, \Phi, \mathcal{A}_1 \ldots, \mathcal{A}_n, \mathcal{K}_1, \ldots, \mathcal{K}_n, I)$, preferences $\mathcal{U}$, endowments $End_1, End_2 \subseteq D$, and a state $\omega^* \in \Omega$ such that $\mathcal{A}_1(\omega^*) = \mathcal{A}_2(\omega^*)$ and $\mathcal{K}_1(\omega^*) \cup \mathcal{K}_2(\omega^*)$ is finite, the following hold:*

(a) If $\langle M, \mathcal{U}\rangle$ satisfies A1 and A2, then there exists a contract that is ω^*-articulable, ω^*-efficient, and ω^*-acceptable for $i = 1, 2$.

(b) If in addition, for $i = 1, 2$, $\mathcal{A}_i$ is constant on $\mathcal{K}_1(\omega^*) \cup \mathcal{K}_2(\omega^*)$, then there exists a contract that is ω^*-articulable, ω'-efficient, and ω'-acceptable for all $\omega' \in \mathcal{K}_1(\omega^*) \cup \mathcal{K}_2(\omega^*)$.

(c) If in addition $\langle M, \mathcal{U}\rangle$ satisfies A3, then there exists a contract $\langle \Lambda, c\rangle$ that is ω^*-articulable, ω'-efficient, and ω'-acceptable for i at all $\omega' \in \mathcal{K}_1(\omega^*) \cup \mathcal{K}_2(\omega^*)$ such that $\Lambda \subseteq \mathcal{L}^{bc}$.

Proof. Define an equivalence relation $\sim$ on states by taking $\omega \sim \omega'$ ff ω and ω' agree on all sentences that the agents are aware of in state ω^* (recall that the agents are aware of the same sentences in ω^*); that is, $\omega \sim \omega'$ if

$$\{\varphi \in \mathcal{L} : (M, \omega) \models \varphi, (M, \omega^*) \models A_1\varphi\} = \{\varphi \in \mathcal{L} : (M, \omega') \models \varphi, (M, \omega^*) \models A_1\varphi\}.$$

(Note that since $\mathcal{A}_1(\omega'^*) = \mathcal{A}_2(\omega^*)$ by assumption, we could replace either occurrence of A_1 above by A_2 without affecting $\sim$.) For each $\omega \in \Omega$, let $[\omega]$ denote the ω's *cell*, that is, the set of states $\sim$-equivalent to ω.

Lemma 7.1. *Let $\kappa : \Omega \rightarrow End_1 \oplus End_2$ be a function such that $\kappa^{-1}((d_1, d_2))$ is a union of $\sim$-cells for all $(d_1, d_2) \in End_1 \oplus End_2$. Then for all finite subsets Ω' of Ω, there exists a ω^*-articulable contract $\langle \Lambda, c\rangle$ such that $c(\omega) = \kappa(\omega)$ for all $\omega \in \Omega'$.*

Proof. Fix a finite subset $\Omega' = \{w_1, \ldots, \omega_n\} \subseteq \Omega$. First suppose that $\omega_i \not\sim \omega_j$ for $i \neq j$, so that for all $i, j \leq n$, $i \neq j$, there exists a sentence $\varphi_{ij} \in \mathcal{L}(\mathcal{A}_1(\omega^*) \cap \mathcal{A}_2(\omega^*))$ such that $(M, \omega_i) \models \varphi_{ij}$ and $(M, \omega_j) \not\models \varphi_{ij}$. For each $i \leq n$ let $\psi_i = \bigwedge_{\{j \neq i\}} \varphi_{i,j}$. For $i = 1, \ldots, n$, let

$$\lambda_i = \bigwedge_{j \neq i} \neg\psi_j \wedge \psi_i;$$

let $\lambda_{n+1} = \wedge_{i=1}^{n} \neg\lambda_i$. Thus, for $i = 1, \ldots, n$, λ_i is true at ω_i and not true at ω_j if $j \neq i$. Moreover, the formulas λ_i are mutually exclusive and exhaustive; exactly one is true at each state in Ω. Finally, let $\Lambda = \{\lambda_i : i \leq n\} \cup \{\bigwedge_{i \leq n} \neg\lambda_i\}$ and define c by taking $c(\lambda_i) = \kappa(\omega_i)$; $c(\bigwedge_{i \leq n} \neg\lambda_i)$ can be defined arbitrarily. Thus, (Λ, c) is a contract such that $c(\omega_i) = \kappa(\omega_i)$ for all $i \leq n$.

Now consider an arbitrary finite subset Ω' of Ω. Let Ω'' be a maximal subset of Ω' that contains at most one element of each $\sim$-cell. We can apply the construction above to Ω'' to get a ω^*-articulable contract (Λ, c) such that $c(\omega) = \kappa(\omega)$ for all $\omega \in \Omega''$. We claim that in fact $c(\omega') = \kappa(\omega')$ for all $\omega' \in \Omega'$. For if $\omega' \in \Omega' - \Omega''$, then $\omega' \sim \omega$ for some $\omega \in \Omega''$. By assumption, $\kappa(\omega') = \kappa(\omega)$, and by construction, we must have $c(\omega') = c(\omega)$. Thus $\kappa(\omega') = c(\omega')$. It follows that $c(\omega') = \kappa(\omega')$ for all $\omega' \in \Omega'$. ☐

Lemma 7.2. *If $\omega, \omega' \in \mathcal{K}_i(w^*)$, $w' \sim w$, and $\mathcal{U}$ satisfies A2, then $U_{i,\omega'} = U_{i,\omega}$.*

Proof. First notice that if $w, w' \in \mathcal{K}_i(w^*)$, then $\mathcal{A}_i(\omega) = \mathcal{A}_i(\omega') = \mathcal{A}_i(\omega^*)$. Thus, by the definition of $\sim$, if $\omega \sim \omega'$, then

$$\{\varphi \in \mathcal{L} : (M, \omega) \models \varphi \wedge A_i\varphi\} = \{\varphi \in \mathcal{L} : (M, \omega') \models \varphi \wedge A_i\varphi\},$$

implying that $\text{PROP}_{\omega}^{\mathcal{A}_i} = \text{PROP}_{\omega'}^{\mathcal{A}_i}$ and $\text{CON}_{\omega}^{\mathcal{A}_i} = \text{CON}_{\omega'}^{\mathcal{A}_i}$. The lemma follows immediately from A2. ☐

For each $\sim$-cell $[\omega]$, define $U_{i,[\omega]} : D \to \mathbb{R}$ by taking $U_{i,[\omega]} = U_{i,\omega'}$ for some $\omega' \in \mathcal{K}_i(\omega^*) \cap [\omega]$ if $\mathcal{K}_i(\omega^*) \cap [\omega]$ is nonempty, and define $U_{i,[\omega]}$ arbitrarily otherwise.

By Lemma 7.2, $U_{i,[\omega]}$ is well defined (in that it does not depend on the choice of ω'). Define

$$\kappa(\omega) \in \arg\max_{(d_1,d_2)\in End_1 \oplus End_2} \{U_{1,[\omega]}(d_1) + U_{2,[\omega]}(d_2) : U_{i,[\omega]}(d_i) \geq \max_{d \in End_i} U_{i,[\omega]}(d), i = 1, 2\}. \quad (1)$$

We assume that the elements of $End_1 \oplus End_2$ are ordered so that if more than element in $End_1 \oplus End_2$ is the argmax in (1), we always choose the least one. This guarantees that $\kappa(\omega) = \kappa(\omega')$ if $\omega \sim \omega'$. Thus, by Lemma 7.1, there exists a ω^*-articulable contract $\hat{c}$ such that $\hat{c}(\omega) = \kappa(\omega)$ for all $\omega \in \mathcal{K}_1(\omega^*) \cup \mathcal{K}_2(\omega^*)$. The next lemma shows that this contract is acceptable and efficient, and thus establishes the claim of part (a).

Lemma 7.3. *The contract $\hat{c}$ is ω^*-efficient and ω^*-acceptable for $i = 1, 2$.*

Proof. For $\omega \in \mathcal{K}_i(\omega^*)$, $U_{i,\omega} = U_{i,[\omega]}$, so ω^*-acceptability follows from the constraint of the maximization problem give by (1).

To see that $\hat{c}$ is ω^*-efficient, suppose, by way of contradiction, that there exists $(d_1, d_2) \in End_1 \oplus End_2$ such that

$$U_{i,\omega^*}(d_i)) \geq U_{i,\omega^*}(\hat{c}_i(\omega^*)) = U_{i,\omega^*}(\kappa_i(\omega^*))$$

for $i = 1, 2$, with at least one inequality strict. Since $\omega^* \in \mathcal{K}_1(\omega^*) \cap \mathcal{K}_2(\omega^*)$, $U_{i,\omega^*} = U_{i,[\omega^*]}$ for $i = 1, 2$. It follows that (d_1, d_2) satisfies the constraints given by (1). But, since one inequality is strict, we have

$$U_{1,[\omega^*]}(d_1) + U_{2,[\omega^*]}(d_2) > U_{1,[\omega^*]}(\kappa_1([\omega^*])) + U_{2,[\omega^*]}((\kappa_2([\omega^*]),$$

contradicting the definition of κ. Hence no such (d_1, d_2) exists, and $\hat{c}$ is ω^*-efficient. □

Part (a) of Theorem 4.1 follows from Lemma 7.3. For part (b), note that if $\mathcal{A}_i$ is constant on $\mathcal{K}_1(\omega^*) \cup \mathcal{K}_2(\omega^*)$ for $i = 1, 2$, Lemma 7.2 can be strengthened to show that if $\omega, \omega' \in \mathcal{K}_1(\omega^*) \cup \mathcal{K}_2(\omega^*)$ then $U_{i,\omega} = U_{i,\omega'}$. The rest of the proof follows follows immediately from the added assumption that $\mathcal{A}_i$ is constant on $\mathcal{K}_1(\omega^*) \cup \mathcal{K}_2(\omega^*)$ for $i = 1, 2$. The rest of the proof follows the proof of part (a) with the obvious adjustments.

For part (c), we consider a coarsening of $\sim$ that identifies states that agree on all sentences $\varphi(d)$ with $\varphi \in \mathcal{L}^{bc}$ that the agents are aware of in ω^*. More precisely, $\omega \sim^{bc} \omega'$ if

$$\{\varphi(d) : \varphi \in \mathcal{L}^{bc}, (M, \omega) \models \varphi(d), (M, \omega^*) \models A_1\varphi(d)\} = \{\varphi(d) : \varphi \in \mathcal{L}^{bc}, (M, \omega') \models \varphi(d), (M, \omega^*) \models A_1\varphi(d)\}.$$

The proof follows that of part (a) except the analogue of Lemma 7.1 uses $\sim^{bc}$-cells and in Lemma 7.2, we appeal to A3 rather than A2. □

8 More on the axiom system

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 φ exactly if i is aware of all the constant, predicate, and concept symbols in φ, 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 φ, i must be aware of φ. $1_{\forall x}$, $1_{\forall X}$, $N_{\forall x}$, $N_{\forall X}$, $K_{\forall x}$, 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 $\forall x K_i(P(x)) \Rightarrow K_i(P(c))$ 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 at ω. It is easy to see that (since agents know what they are aware of) $\forall X K_i A_i(X(c))$ holds at ω. However, if i considers a world ω' such that $\Phi(\omega')$ includes Q and i is not aware of Q at ω', then $K_i(\forall X A_i(X(c)))$ does not hold at ω. 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 φ is a formula with a free predicate variable X. Then Barcan$_X$ says that if, in a state ω, agent i is aware of all the predicates in φ and if, no matter what formula $\psi \in \mathcal{L}^{bc} \cap \Phi(\omega)$ is substituted for X in φ, if i is aware of all the predicates in ψ, then i knows $\varphi[X/\psi]$, then in any state ω' 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 sentence, 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$_X$ (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.

Although MP is standard, it is not obviously sound. To understand the problem, suppose that φ and ψ are sentences such that φ and $\varphi \Rightarrow \psi$ are valid. To show that MP is sound, we would have to show that φ is valid. Suppose not. Then there is a model M^* and state ω^* in M^* such that $(M^*, \omega^*) \models \neg\psi$. If $(M^*, \omega^*) \models \varphi$ and $(M^*, \omega^*) \models \varphi \Rightarrow \psi$, then we have an immediate contradiction. Unfortunately, despite the validity of φ and $\varphi \Rightarrow \psi$, we cannot conclude that $(M^*, \omega^*) \models \varphi$ and that $(M^*, \omega^*) \models \varphi \Rightarrow \psi$. These sentences might not be in the language at state ω. The assumption that $\mathscr{P}$ is infinite allows us to deal with this difficulty.

We now prove Theorem 5.1, which says that AX is a sound and complete axiomatization if the set $\mathscr{P}$ of predicates is infinite. Since the proof follows exactly the same lines as the HR soundness and completeness proof, we focus here on the main differences.

Soundness is straightforward, with the exception of MP, Gen$_{\forall_x}$, Gen$_{\forall_X}$, and Barcan$_X$. The soundness of MP is proved in Lemma A.1, Proposition A.1, and Corollaries A.1 and A.2 in (Halpern and Rêgo 2013), under the assumption that the set of primitive propositions is infinite; the soundness of Barcan$_X$ is proved in Proposition A.3 in (Halpern and Rêgo 2013) (where the axiom is labeled Barcan$^*_X$). Essentially the identical proofs show soundness in our setting. We now prove that Gen$_{\forall_x}$ and Gen$_{\forall_X}$ are sound.

Proposition 8.1. *Gen$_{\forall_x}$ and Gen$_{\forall_X}$ are sound.*

Proof. For Gen$_{\forall_x}$, suppose that $\models \varphi$. We claim that we must have $\models \varphi[c/c']$ for all constants c'. For suppose not. Then there must be some model M and state ω such that $(M, \omega) \models \neg\varphi[c/c']$. Let M' be a model just like M, but with the roles of c and c' reversed. Specifically, if I^M and $I^{M'}$ denote the interpretation functions of M and M', respectively,[9] then for each predicate P and each state ω, $I^{M'}$ is the result of replacing c in P_ω by c' and replacing c' in P_ω by c. It is easy to show, by induction on the structure of ψ, that, for all sentences ψ, we have that $(M, \omega) \models \psi[c/c']$ iff $(M', \omega) \models \psi$. (In doing the induction, we first consider how many occurrences of $\forall X$, that is, how many instances of quantification over predicates, there are in ψ, and then do a subinduction on the length of ψ.) It follows that $(M', \omega) \models \neg\varphi$, contradicting the validity of φ. Thus, $\models \varphi[c/c']$ for all constants c', so $\models \forall x\varphi$, as desired.

For Gen$_{\forall_X}$, again suppose that $\models \varphi$. We want to show that $\models \forall X\varphi[P/X]$. Suppose not. Then there must be some model M, state ω in M, and sentence $\psi \in \mathcal{L}^{bc}(\Phi^M(\omega))$ such that $(M, \omega) \models \neg\varphi[P/\psi]$. Let M' be identical to M except that, for each state ω', (a) $\Phi^{M'}(\omega')$ is the resulting of adding P to $\Phi^{M'}(\omega')$ iff $\psi \in \mathcal{L}(\Phi^M(\omega'))$; (b) $I^{M'}$ is such that, for all $c \in \mathscr{O}$, we have that $(M', \omega') \models P(c)$ iff $(M, \omega') \models \psi(c)$; and (c) the awareness functions are such that, for all agents i, $(M', \omega') \models A_i(P(c))$ iff $(M, \omega') \models A_i(\psi(c))$. Intuitively, ψ plays the same role in M that P does in M'.

Then a straightforward induction on the structure of sentences shows that for all states ω' and all sentences φ', we have that $(M, \omega') \models \varphi'[P/\psi]$ iff $(M', \omega') \models \varphi'$. Since, by assumption, $(M, \omega) \models \neg\varphi[P/\psi]$, it follows $(M', \omega') \models \neg\varphi$, contradicting the validity of φ. Hence, we have that $\forall X\varphi[P/X]$, as desired. □

For completeness, fix a language $\mathcal{L}(\mathscr{O}, \mathscr{P}, \mathscr{C})$ where $\mathscr{P}$ is infinite. We now list the major lemmas used by HR to prove completeness, expanding only on the differences. As usual, the idea of the completeness proof is to construct a canonical model M^c where the worlds are maximal consistent sets of sentences. We then show that if ω_Γ is the world corresponding to the maximal consistent set Γ, then $(M^c, \omega_\Gamma) \models \varphi$ iff $\varphi \in \Gamma$. As observed in (Halpern and Rêgo 2009), this will not quite work in the presence of quantification; there may be a maximal consistent set Γ of sentences such that $\neg\forall X\varphi \in \Gamma$, but $\varphi[X/\psi] \in \Gamma$ for all $\psi \in \mathcal{L}^{bc}$. That is, there is no witness to the falsity of $\forall X\varphi$ in Γ. This problem was dealt with in (Halpern and Rêgo 2009) by restricting to maximal consistent sets Γ that are *acceptable* in the sense that if $\neg\forall X\varphi \in \Gamma$, then $\neg\varphi[X/q] \in \Gamma$ for infinitely many primitive propositions $q \in \Phi$. (Recall that HR consider a propositional language; this notion of acceptability requires the language to include infinitely many primitive propositions, as HR assumed it did.) As in (Halpern and Rêgo 2013), because here we have possibly different languages associated different worlds, we need to consider acceptability and maximality with respect to a language. The following definition is adapted from (Halpern and Rêgo 2013).

Definition 8.1. *A sentence φ is* provable *from AX, denoted AX $\vdash \varphi$, if there is a sequence of sentences such that the last one is φ, and each one is either an instance of an axiom of AX or follows from previous sentences in the sequence by an application of an inference rule. We write $\Gamma \vdash \varphi$ if there are sentences $\beta_1, \ldots, \beta_m$ in Γ such that AX $\vdash (\beta_1 \wedge \ldots \wedge \beta_m) \Rightarrow \varphi$.*

Definition 8.2. *A set Γ of sentences is* acceptable with respect to the language *$\mathcal{L}(\mathscr{O}, \mathscr{P}', \mathscr{C}')$ if, for all sentences $\varphi \in \mathcal{L}(\mathscr{O}, \mathscr{P}', \mathscr{C}')$, if $\Gamma \vdash \varphi[X/P]$ for all but finitely predicate symbols $P \in \mathscr{P}'$, then $\Gamma \vdash \forall X\varphi$.*

Definition 8.3. *Γ is a* maximal AX-consistent set of sentences *with respect to $\mathcal{L}(\mathscr{O}, \mathscr{P}', \mathscr{C}') \subseteq \mathcal{L}(\mathscr{O}, \mathscr{P}, \mathscr{C})$ if Γ is an AX-consistent subset of $\mathcal{L}(\mathscr{O}, \mathscr{P}', \mathscr{C}')$ and, for all sentences $\varphi \in \mathcal{L}(\mathscr{O}, \mathscr{P}', \mathscr{C}')$, if $\Gamma \cup \{\varphi\}$ is AX-consistent, then $\varphi \in \Gamma$.*

The following four lemmas are essentially Lemmas A.4, A.5, A.6, and A.7 in (Halpern and Rêgo 2009). Since the proofs are essentially identical, we do not repeat them here.

Lemma 8.1. *If Γ is a finite set of sentences, then Γ is acceptable with respect to every language $L(\mathscr{O}, \mathscr{P}', \mathscr{C}')$ such that $\mathscr{P}'$ is infinite.*

Lemma 8.2. *If Γ is acceptable with respect to $\mathcal{L}(\mathscr{O}, \mathscr{P}', \mathscr{C}')$ and ψ is a sentence in $\mathcal{L}(\mathscr{O}, \mathscr{P}', \mathscr{C}')$, then $\Gamma \cup \{\psi\}$ is acceptable with respect to $\mathcal{L}(\mathscr{O}, \mathscr{P}', \mathscr{C}')$.*

Lemma 8.3. *If $\Gamma \subseteq \mathcal{L}(\mathscr{O}, \mathscr{P}', \mathscr{C}')$ is an acceptable AX-consistent set of sentences with respect to $\mathcal{L}(\mathscr{O}, \mathscr{P}', \mathscr{C}')$, then Γ can be extended to a set of sentences that is acceptable and maximal AX-consistent with respect to $\mathcal{L}(\mathscr{O}', \mathscr{P}', \mathscr{C}')$.*

Let $\Gamma/K_i = \{\varphi : K_i\varphi \in \Gamma\}$.

Lemma 8.4. *If Γ is a maximal AX-consistent set of sentences with respect to $\mathcal{L}(\mathscr{O}, \mathscr{P}', \mathscr{C}')$ that contains $\neg K_i\varphi$ and $A_i\varphi$, then $\Gamma/K_i \cup \{\neg\varphi\}$ is AX-consistent.*

The next lemma is essentially Lemma A.6 in (Halpern and Rêgo 2013), and again, its proof is almost identical. We say $\mathscr{P}'$ is *co-infinite* if $\mathscr{P} - \mathscr{P}'$ is infinite.

[9] We use analogous notation consistently below.

Lemma 8.5. *If Γ is an acceptable maximal AX-consistent set of sentences with respect to $\mathcal{L}(\mathscr{O}, \mathscr{P}', \mathscr{C}')$, where $\mathscr{P}'$ is infinite and co-infinite, $\neg K_i\varphi \in \Gamma$, and $A_i\varphi \in \Gamma$, then there is an infinite and co-infinite set $\mathscr{P}''$ of predicates and a set Δ of sentences that is an acceptable, maximal AX-consistent set with respect to $\mathcal{L}(\mathscr{O}, \mathscr{P}'', \mathscr{C}')$ and contains $\Gamma/K_i \cup \{\neg\varphi\}$. Moreover, $A_i\psi \in \Delta$ iff $A_i\psi \in \Gamma$ for all sentences ψ.*

The following lemma is the one new lemma that we need to deal with concepts and variables. It illustrates the role of acceptability in the construction.

Lemma 8.6. *If Γ is an acceptable maximal AX-consistent set of sentences with respect to $\mathcal{L}(\mathscr{O}, \mathscr{P}', \mathscr{C}')$ and $C \in \mathscr{C}'$, then there exists a sentence $\psi \in \mathcal{L}^{bc}(\mathscr{P}')$ such that $\forall x(C(x) \Leftrightarrow \psi(x)) \in \Gamma$.*

Proof. Suppose, by way of contradiction, that there is no sentence $\psi \in \mathcal{L}^{bc}(\mathscr{P}')$ such that $\forall x(C(x) \Leftrightarrow \psi(x)) \in \Gamma$. Then, since Γ is maximal, for all predicates $Q \in \mathscr{P}'$, we must have that $\neg\forall x(C(x) \Leftrightarrow Q(x)) \in \Gamma$. Since Γ is acceptable, it follows that $\forall X\neg\forall x(C(x) \Leftrightarrow X(x)) \in \Gamma$. But since Γ contains every instance of axiom Con, it contains $\exists X(\forall x(C(x) \Leftrightarrow X(x)))$ (i.e., $\neg\forall X\neg(\forall x(C(x) \Leftrightarrow X(x))) \in \Gamma$). Thus, Γ is inconsistent, a contradiction. □

We now complete the completeness proof by constructing a canonical model, essentially as is done in (Halpern and Rêgo 2013).

Lemma 8.7. *If φ is an AX-consistent sentence, then φ is satisfiable.*

Proof. We construct a canonical model where the worlds are maximal consistent sets of sentences. However, now the worlds must also explicitly include the language, and we assume that the language associated with each world is infinite and co-infinite. Let $M^c = (\Omega^c, D, \Phi^c, \mathcal{A}_1^c, \ldots, \mathcal{A}_n^c, \mathcal{K}_1^c, \ldots, \mathcal{K}_n^c, I^c)$ be the canonical awareness structure for $\mathcal{L}(\mathscr{O}, \mathscr{P}, \mathscr{C})$, constructed as follows:

- $\Omega^c = \{(\omega_\Gamma, \mathcal{L}(\mathscr{O}, \mathscr{P}', \mathscr{C}')) :$ Γ is a set of sentences that is acceptable and maximal AX-consistent with respect to $\mathcal{L}(\mathscr{O}, \mathscr{P}', \mathscr{C}')$, $L(\mathscr{O}, \mathscr{P}', \mathscr{C}') \subseteq \mathcal{L}(\mathscr{O}, \mathscr{P}, \mathscr{C})$, and $\mathscr{P}'$ is infinite and co-infinite$\}$;
- $D = \mathscr{O}$;
- $\Phi^c((\omega_\Gamma, \mathcal{L})) = \mathcal{L}$;
- $\mathcal{A}_i^c \omega_\Gamma, \mathcal{L}) = \{\psi : A_i\psi \in \Gamma\}$;
- $\mathcal{K}_i^c((\omega_\Gamma, \mathcal{L})) = \{(\omega_\Delta, \mathcal{L}') : \Gamma/X_i \subseteq \Delta$ and $A_i\varphi \in \Gamma$ iff $A_i\varphi \in \Delta$ for all sentences $\varphi\}$;
- $I^c_{(\omega_\Gamma, \mathcal{L})}(P) = \{c : P(c) \in \Gamma\}$;
- $I^c_{(\omega_\Gamma, \mathcal{L})}(C) = \psi$ if $\forall x(C(x) \Leftrightarrow \psi(x)) \in \Gamma$. (By Lemma 8.6, there will be such a ψ; if there is one more than one, we can pick the least one in some ordering of formulas.)

Standard arguments now show that $\psi \in \Gamma$ iff $(M^c, (\omega_\Gamma, \mathcal{L})) \models \psi$. It follows from Lemmas A.1 and A.3 that every consistent sentence is in some acceptable and maximal consistent set with respect to a language $\mathcal{L}(O, \mathscr{P}', \mathscr{C}')$ such that $\mathscr{P}'$ is infinite and co-infinite. Thus, if φ is consistent, there is some state $(\omega_\Gamma, \mathcal{L})$ in the canonical model such that $(M^c, (\omega_\Gamma, \mathcal{L})) \models \varphi$. This proves the lemma and the theorem. □