Reasoning About Knowledge of Unawareness Revisited

Joseph Y. Halpern
Computer Science Department
Cornell University
Ithaca, NY, 14853, U.S.A.
halpern@cs.cornell.edu

Leandro C. Rêgo
Statistics Department
Federal University of Pernambuco
Recife, PE, 50740-040, Brazil
leandro@de.ufpe.br

Abstract

In earlier work [Halpern and Rêgo 2006b], we proposed a logic that extends the Logic of General Awareness of Fagin and Halpern [1988] by allowing quantification over primitive propositions. This makes it possible to express the fact that an agent knows that there are some facts of which he is unaware. In that logic, it is not possible to model an agent who is uncertain about whether he is aware of all formulas. To overcome this problem, we keep the syntax of the earlier paper, but allow models where, with each world, a possibly different language is associated. We provide a sound and complete axiomatization for this logic and show that, under natural assumptions, the quantifier-free fragment of the logic is characterized by exactly the same axioms as the logic of Heifetz, Meier, and Schipper [2008].

1 INTRODUCTION

Adding awareness to standard models of epistemic logic has been shown to be useful in describing many situations (see [Fagin and Halpern 1988; Heifetz, Meier, and Schipper 2006] for some examples). One of the best-known models of awareness is due to Fagin and Halpern [1988] (FH from now on). They add an awareness operator to the language, and associate with each world in a standard possible-worlds model of knowledge a set of formulas that each agent is aware of. They then say that an agent *explicitly knows* a formula φ if φ is true in all worlds that the agent considers possible (the traditional definition of knowledge, going back to Hintikka [1962]) and the agent is aware of φ.

In the economics literature, going back to the work of Modica and Rustichini [1994, 1999] (MR from now on), a somewhat different approach is taken. A possibly different set $\mathcal{L}(s)$ of primitive propositions is associated with each world s. Intuitively, at world s, the agent is aware only of formulas that use the primitive propositions in $\mathcal{L}(s)$. A definition of knowledge is given in this framework, and the agent is said to be aware of φ if, by definition, $K_i\varphi \vee K_i\neg K_i\varphi$ holds. Heifetz, Meier, and Schipper [2006, 2008] (HMS from now on), extend the ideas of MR to a multiagent setting. This extension is nontrivial, requiring lattices of state spaces, with projection functions between them. As we showed in earlier work [Halpern 2001; Halpern and Rêgo 2008], the work of MR and HMS can be seen as a special case of the FH approach, where two assumptions are made on awareness: awareness is *generated by primitive propositions*, that is, an agent is aware of a formula iff he is aware of all primitive propositions occurring in it, and agents know what they are aware of (so that they are aware of the same formulas in all worlds that they consider possible).

As we pointed out in [Halpern and Rêgo 2006b] (referred to as HR from now on), if awareness is generated by primitive propositions, then it is impossible for an agent to (explicitly) know that he is unaware of a specific fact. Nevertheless, an agent may well be aware that there are relevant facts that he is unaware of. For example, primary-care physicians know that specialists are aware of things that could improve a patient's treatment that they are not aware of; investors know that investment fund companies may be aware of issues involving the financial market that could result in higher profits that they are not aware of. It thus becomes of interest to model knowledge of lack of awareness. HR does this by extending the syntax of the FH approach to allow quantification, making it possible to say that an agent knows that there exists a formula of which the agent is unaware. A complete axiomatization is provided for the resulting logic. Unfortunately, the logic has a significant problem if we assume the standard properties of knowledge and awareness: it is

TARK '09, July 6-8, 2009, California
ISBN: 978-1-60558-560-4...$10.00

impossible for an agent to be uncertain about whether he is aware of all formulas.

In this paper, we deal with this problem by considering the same language as in HR (so that we can express the fact that an agent knows that he is not aware of all formulas, using quantification), but using the idea of MR that there is a different language associated with each world. As we show, this slight change makes it possible for an agent to be uncertain about whether he is aware of all formulas, while still being aware of exactly the same formulas in all worlds he considers possible. We provide a natural complete axiomatization for the resulting logic. Interestingly, knowledge in this logic acts much like explicit knowledge in the original FH framework, if we take "awareness of φ" to mean $K_i(\varphi \vee \neg\varphi)$; intuitively, this is true if all the primitive propositions in φ are part of the language at all worlds that i considers possible. Under minimal assumptions, $K_i(\varphi \vee \neg\varphi)$ is shown to be equivalent to $K_i\varphi \vee K_i\neg K_i\varphi$: in fact, the quantifier-free fragment of the logic that just uses the K_i operator is shown to be characterized by exactly the same axioms as the HMS approach, and awareness can be defined the same way. Thus, we can capture the essence of MR and HMS approach using simple semantics and being able to reason about knowledge of lack of awareness.

Board and Chung [2009] independently pointed out the problem of the HR model and proposed the solution of allowing different languages at different worlds. They also consider a model of awareness with quantification, but they use first-order modal logic, so their quantification is over domain elements. Moreover, they take awareness with respect to domain elements, not formulas; that is, agents are (un)aware of objects (i.e., domain elements), not formulas. They also allow different domains at different worlds; more precisely, they allow an agent to have a subjective view of what the set of objects is at each world. Sillari [2008] uses much the same approach as Board and Chung [2009]. That is, he has a first-order logic of awareness, where the quantification and awareness is with respect to domain elements, and also allows from different subjective domains at each world.

The rest of the paper is organized as follows. In Section 2, we review the HR model of knowledge of unawareness. In Section 3, we present our new logic and axiomatize it in Section 4. In Section 5, we compare our logic with that of HMS and discuss awareness more generally. All proofs are left to the full paper, which can be found at www.cs.cornell.edu/home/halpern/papers/tark09.pdf.

2 THE HR MODEL

In this section, we briefly review the relevant results of [Halpern and Rêgo 2006b]. The syntax of the logic is as follows: given a set $\{1,\ldots,n\}$ of agents, formulas are formed by starting with a countable set $\Phi = \{p, q, \ldots\}$ of primitive propositions and a countable set $\mathcal{X}$ of variables, and then closing off under conjunction ($\wedge$), negation ($\neg$), the modal operators K_i, A_i, X_i, $i = 1,\ldots,n$. We also allow for quantification over variables, so that if φ is a formula, then so is $\forall x\varphi$. Let $\mathcal{L}_n^{\forall,K,X,A}(\Phi,\mathcal{X})$ denote this language and let $\mathcal{L}_n^{K,X,A}(\Phi)$ be the subset of formulas that do not mention quantification or variables. As usual, we define $\varphi \vee \psi$, $\varphi \Rightarrow \psi$, and $\exists x\varphi$ as abbreviations of $\neg(\neg\varphi \wedge \neg\psi)$, $\neg\varphi \vee \psi$, and $\neg\forall x\neg\varphi$, respectively. The intended interpretation of $A_i\varphi$ is "i is aware of φ".

Essentially as in first-order logic, we can define inductively what it means for a variable x to be *free* in a formula φ. Intuitively, an occurrence of a variable is free in a formula if it is not bound by a quantifier. A formula that contains no free variables is called a *sentence*. We are ultimately interested in sentences. If ψ is a formula, let $\varphi[x/\psi]$ denote the formula that results by replacing all free occurrences of the variable x in φ by ψ. (If there is no free occurrence of x in φ, then $\varphi[x/\psi] = \varphi$.) In quantified modal logic, the quantifiers are typically taken to range over propositions (intuitively, sets of worlds), but this does not work in our setting because awareness is syntactic; when we write, for example, $\forall xA_ix$, we essentially mean that $A_i\varphi$ holds for all *formulas* φ. However, there is another subtlety. If we define $\forall x\varphi$ to be true if $\varphi[x/\psi]$ is true for *all* formulas ψ, then there are problems giving semantics to a formula such as $\varphi = \forall x(x)$, since $\varphi[x/\varphi] = \varphi$. We avoid these difficulties by taking the quantification to be over quantifier-free sentences. (See [Halpern and Rêgo 2006b] for further discussion.)

We give semantics to sentences in $\mathcal{L}_n^{\forall,K,X,A}(\Phi,\mathcal{X})$ in awareness structures. A tuple $M = (S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{A}_1, \ldots, \mathcal{A}_n)$ is an *awareness structure for n agents (over Φ)* if S is a set of worlds, $\pi : S \times \Phi \rightarrow \{\mathbf{true}, \mathbf{false}\}$ is an interpretation that determines which primitive propositions are true at each world, $\mathcal{K}_i$ is a binary relation on S for each agent $i = 1,\ldots,n$, and $\mathcal{A}_i$ is a function associating a set of sentences with each world in S, for $i = 1,...,n$. Intuitively, if $(s,t) \in \mathcal{K}_i$, then agent i considers world t possible at world s, while $\mathcal{A}_i(s)$ is the set of sentences that agent i is aware of at world s. We are often interested in awareness structures where the $\mathcal{K}_i$ relations satisfy some properties of interest, such as reflexivity, transitivity, or the *Euclidean* property (if $(s,t),(s,u) \in \mathcal{K}_i$, then $(t,u) \in \mathcal{K}_i$). It is well known that these properties of the relation correspond to properties of knowledge of interest (see Theorem 2.1 and the following discussion). We often abuse notation and define $\mathcal{K}_i(s) = \{t : (s,t) \in \mathcal{K}_i\}$, thus writing $t \in \mathcal{K}_i(s)$ rather than $(s,t) \in \mathcal{K}_i$. This notation allows us to view a binary relation $\mathcal{K}_i$ on S as a *possibility correspondence*, that is, a function from S to 2^S. (The use of possibility correspondences is more standard in the economics literature than binary relations, but they are clearly essentially equivalent.)

Semantics is given to sentences in $\mathcal{L}_n^{\forall,K,X,A}(\Phi,\mathcal{X})$ by induction on the number of quantifiers, with a subinduction on the length of the sentence. Truth for primitive propositions, for $\neg$, and for $\wedge$ is defined in the usual way. The other cases are defined as follows:[1]

$(M,s) \models K_i\varphi$ if $(M,t) \models \varphi$ for all $t \in \mathcal{K}_i(s)$
$(M,s) \models A_i\varphi$ if $\varphi \in \mathcal{A}_i(s)$
$(M,s) \models X_i\varphi$ if $(M,s) \models A_i\varphi$ and $(M,s) \models K_i\varphi$
$(M,s) \models \forall x\varphi$ if $(M,s) \models \varphi[x/\psi], \forall\psi \in \mathcal{L}_n^{K,X,A}(\Phi)$.

There are two standard restrictions on agents' awareness that capture the assumptions typically made in the game-theoretic literature [Modica and Rustichini 1999; Heifetz, Meier, and Schipper 2006; Heifetz, Meier, and Schipper 2008]. We describe these here in terms of the awareness function, and then characterize them axiomatically.

- Awareness is *generated by primitive propositions (agpp)* if, for all agents i, $\varphi \in \mathcal{A}_i(s)$ iff all the primitive propositions that appear in φ are in $\mathcal{A}_i(s) \cap \Phi$.
- *Agents know what they are aware of (ka)* if, for all agents i and all worlds s,t such that $(s,t) \in \mathcal{K}_i$ we have that $\mathcal{A}_i(s) = \mathcal{A}_i(t)$.

For ease of exposition, we restrict in this paper to structures that satisfy *agpp* and *ka*. If C is a (possibly empty) subset of $\{r,t,e\}$, then $\mathcal{M}_n^C(\Phi,\mathcal{X})$ is the set of all awareness structures such that awareness satisfies *agpp* and *ka* and the possibility correspondence is reflexive (r), transitive (t), and Euclidean (e) if these properties are in C.

A sentence $\varphi \in \mathcal{L}_n^{\forall,K,X,A}(\Phi,\mathcal{X})$ is said to be *valid* in awareness structure M, written $M \models \varphi$, if $(M,s) \not\models \neg\varphi$ for all $s \in S$. (This notion is called *weak validity* in [Halpern and Rêgo 2008]. For the semantics we are considering here, weak validity is equivalent to the standard notion of validity, where a formula is valid in an awareness structure if it is true at all worlds in that structure. However, in the next section, we modify the semantics to allow some formulas to be undefined at some worlds; with this change, the two notions do not coincide. As we use weak validity in the next section, we use the same definition here for the sake of uniformity.) A sentence is valid in a class $\mathcal{M}$ of awareness structures, written $\mathcal{M} \models \varphi$, if it is valid for all awareness structures in $\mathcal{M}$, that is, if $M \models \varphi$ for all $M \in \mathcal{M}$.

In [Halpern and Rêgo 2006b], we gave sound and complete axiomatizations for both the language $\mathcal{L}_n^{\forall,K,X,A}(\Phi,\mathcal{X})$ and the language $\mathcal{L}_n^{\forall,X,A}(\Phi,\mathcal{X})$, which does not mention the implicit knowledge operator K_i (and the quantification is thus only over sentences in $\mathcal{L}_n^{X,A}(\Phi)$). The latter language is arguably more natural (since agents do not have access to the implicit knowledge modeled by K_i), but some issues become clearer when considering both. We start by describing axioms for the language $\mathcal{L}_n^{\forall,K,X,A}(\Phi,\mathcal{X})$, and then describe how they are modified to deal with $\mathcal{L}_n^{\forall,X,A}(\Phi,\mathcal{X})$. Given a formula φ, let $\Phi(\varphi)$ be the set of primitive propositions in Φ that occur in φ.

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

AGPP. $A_i\varphi \Leftrightarrow \wedge_{p\in\Phi(\varphi)} A_i p$.[2]

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

NKA. $\neg A_i\varphi \Rightarrow K_i\neg A_i\varphi$

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

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

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

5. $\neg K_i\varphi \Rightarrow K_i\neg K_i\varphi$.

A0. $X_i\varphi \Leftrightarrow K_i\varphi \wedge A_i\varphi$.

$1_\forall$. $\forall x\varphi \Rightarrow \varphi[x/\psi]$ if ψ is a quantifier-free sentence.

$K_\forall$. $\forall x(\varphi \Rightarrow \psi) \Rightarrow (\forall x\varphi \Rightarrow \forall x\psi)$.

$N_\forall$. $\varphi \Rightarrow \forall x\varphi$ if x is not free in φ.

[1] HR gives semantics to arbitrary formulas, including formulas with free variables. This requires using *valuations* that give meaning to free variables. By restricting to sentences, which is all we are ultimately interested in, we are able to dispense with valuations here, and thus simplify the presentation of the semantics.

[2] As usual, the empty conjunction is taken to be the vacuously true formula *true*, so that $A_i\varphi$ is vacuously true if no primitive propositions occur in φ. We remark that in the conference version of HR, an apparently weaker version of AGPP called *weak generation of awareness by primitive propositions* is used. However, this is shown in HR to be equivalent to AGPP if the agent is aware of at least one primitive proposition, so AGPP is used in the final version of HR, and we use it here as well.

Barcan. $\forall x K_i \varphi \Rightarrow K_i \forall x \varphi$.

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

Gen_K. From φ infer $K_i \varphi$.

$\text{Gen}_\forall$. If q is a primitive proposition, then from φ infer $\forall x \varphi[q/x]$.

Axioms Prop, K, T, 4, 5 and inference rules MP and Gen_K are standard in epistemic logics. A0 captures the relationship between explicit knowledge, implicit knowledge and awareness. Axioms $1_\forall$, $K_\forall$, $N_\forall$ and inference rules $\text{Gen}_\forall$ are standard for propositional quantification.[3] The Barcan axiom, which is well-known in first-order modal logic, captures the relationship between quantification and K_i. Axioms AGPP, KA, and NKA capture the properties of awareness being generated by primitive propositions and agents knowing which formulas they are aware of. Let $\text{AX}^{\text{K,X,A},\forall}$ be the axiom system consisting of all the axioms and inference rules in {Prop, AGPP, KA, NKA, K, A0, $1_\forall$, $K_\forall$, $N_\forall$, Barcan, MP, Gen_K, $\text{Gen}_\forall$}.

The language $\mathcal{L}_n^{\forall,X,A}$ without the modal operators K_i has an axiomatization that is similar in spirit. Let K_X, T_X, 4_X, XA, and Barcan_X be the axioms that result by replacing the K_i in K, T, 4, KA, and Barcan, respectively, by X_i. Let 5_X and Gen_X be the axioms that result from adding awareness to 5 and Gen_K:

5_X. $(\neg X_i \varphi \wedge A_i \varphi) \Rightarrow X_i \neg X_i \varphi$.

Gen_X. From φ infer $A_i \varphi \Rightarrow X_i \varphi$.

The analogue of axiom NKA written in terms of X_i, $\neg A_i \varphi \Rightarrow X_i \neg A_i \varphi$, is not valid. To get completeness in models where agents know what they are aware of, we need the following axiom, which can be viewed as a weakening of NKA:

FA_X. $\neg \forall x A_i x \Rightarrow X_i \neg \forall x A_i x$.

Finally, consider the following axiom that captures the relation between explicit knowledge and awareness:

A0_X. $X_i \varphi \Rightarrow A_i \varphi$.

Let $\text{AX}^{\text{X,A},\forall}$ be the axiom system consisting of all the the axioms and inference rules in {Prop, AGPP, XA, FA_X, K_X, A0_X, $1_\forall$, $K_\forall$, $N_\forall$, Barcan_X, MP, Gen_X, $\text{Gen}_\forall$}. The following result shows that the semantic properties r, t, e are captured by the axioms T, 4, and 5, respectively in the language $\mathcal{L}_n^{\forall,K,X,A}$; similarly, these same properties are captured by T_X, 4_X, and 5_X in the language $\mathcal{L}_n^{\forall,X,A}$.

Theorem 2.1: [Halpern and Rêgo 2006b] *If $\mathcal{C}$ (resp., $\mathcal{C}_X$) is a (possibly empty) subset of* $\{\text{T}, 4, 5\}$ *(resp.,* $\{\text{T}_\text{X}, 4_\text{X}, 5_\text{X}\}$*) and if C is the corresponding subset of* $\{r, t, e\}$ *then* $\text{AX}^{\text{K,X,A},\forall} \cup \mathcal{C}$ *(resp.,* $\text{AX}^{\text{X,A},\forall} \cup \mathcal{C}_\text{X}$*) is a sound and complete axiomatization of the sentences in* $\mathcal{L}_n^{\forall,K,X,A}(\Phi, \mathcal{X})$ *(resp.* $\mathcal{L}_n^{\forall,X,A}(\Phi, \mathcal{X})$*) with respect to* $\mathcal{M}_n^C(\Phi, \mathcal{X})$.

Consider the formula $\psi = \neg X_i \neg \forall x A_i x \wedge \neg X_i \forall x A_i x$. The formula ψ says that agent i considers it possible that she is aware of all formulas and also considers it possible that she is not aware of all formulas. It is not hard to show ψ is not satisfiable in any structure in $\mathcal{M}(\Phi, \mathcal{X})$, so $\neg\psi$ is valid in awareness structures in $\mathcal{M}(\Phi, \mathcal{X})$, It seems reasonable that an agent can be uncertain about whether there are formulas he is unaware of. In the next section, we show that a slight modification of the HR approach using ideas of MR, allows this, while still maintaining the desirable properties of the HR approach.

3 THE NEW MODEL

We keep the syntax of Section 2, but, following MR, we allow different languages to be associated with different worlds. Define an *extended awareness structure for n agents (over Φ)* to be a tuple $M = (S, \mathcal{L}, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{A}_1, \ldots, \mathcal{A}_n)$, where $M = (S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{A}_1, \ldots, \mathcal{A}_n)$ is an awareness structure and $\mathcal{L}$ maps worlds in S to nonempty subsets of Φ. Intuitively, $\mathcal{L}_n^{\forall,K,X,A}(\mathcal{L}(s), \mathcal{X})$ is the language associated with world s. We require that $\mathcal{A}_i(s) \subseteq \mathcal{L}_n^{\forall,K,X,A}(\mathcal{L}(s), \mathcal{X})$, so that an agent can be aware only of sentences that are in the language of the current world. We still want to require that *agpp* and *ka*; this means that if $(s,t) \in \mathcal{K}_i$, then $\mathcal{A}_i(s) \subseteq \mathcal{L}_n^{\forall,K,X,A}(\mathcal{L}(t), \mathcal{X})$. But $\mathcal{L}(t)$ may well include primitive propositions that the agent is not aware of at s. It may at first seem strange that an agent considers possible a world whose language includes formulas of which he is not aware. (Note that, in general, this happens in the HR approach too, even though there we require that $\mathcal{L}(s) = \mathcal{L}(t)$.) But, in the context of knowledge of lack awareness, there is an easy explanation for this: the fact that $\mathcal{A}_i(s)$ is a strict subset of the sentences in $\mathcal{L}_n^{\forall,K,X,A}(\mathcal{L}(t), \mathcal{X})$ is just our way of modeling that the agent considers it possible that there are formulas of which he is unaware; he can even "name" or "label" these formulas, although he may not understand what the names re-

[3] Since we gave semantics not just to sentences, but also to formulas with free variables in [Halpern and Rêgo 2006b], we were able to use a simpler version of $\text{Gen}_\forall$ that applies to arbitrary formulas: from φ infer $\forall x \varphi$. Note that all the other axioms and inference rules apply without change to formulas as well as sentences.

fer to. If the agent considers possible a world t where $\mathcal{A}_i(s)$ consists of every sentence in $\mathcal{L}_n^{\forall,K,X,A}(\mathcal{L}(t),\mathcal{X})$, then the agent considers it possible that he is aware of all formulas. The formula ψ in Section 2 is satisfied at a world s where agent i considers possible a world t_1 such that $\mathcal{A}_i(s)$ consists of all sentences in $\mathcal{L}_n^{\forall,K,X,A}(\mathcal{L}(t_1),\mathcal{X})$ and a world t_2 such that $\mathcal{A}_i(s)$ does not contain some sentence in $\mathcal{L}_n^{\forall,K,X,A}(\mathcal{L}(t_2),\mathcal{X})$. Note that we can also describe worlds where agent 1 considers it possible that agents 2 and 3 are aware of the same formulas, although both are aware of formulas that he (1) is not aware of, and other more complicated relationships between the awareness of agents. See Section 5 for further discussion of awareness of unawareness in this setting.

The truth relation is defined for formulas in $\mathcal{L}_n^{\forall,K,X,A}(\Phi,\mathcal{X})$ just as in Section 2, except that for a formula φ to be true at a world s, we also require that $\varphi \in \mathcal{L}_n^{\forall,K,X,A}(\mathcal{L}(s),\mathcal{X})$, so we just add this condition everywhere. Thus, for example,

- $(M,s) \models p$ if $p \in \mathcal{L}(s)$ and $\pi(s,p) = \mathbf{true}$;
- $(M,s) \models \neg\varphi$ if $\varphi \in \mathcal{L}_n^{\forall,K,X,A}(\mathcal{L}(s),\mathcal{X})$ and $(M,s) \not\models \varphi$.
- $(M,s) \models \forall x\varphi$ if $\varphi \in \mathcal{L}_n^{\forall,K,X,A}(\mathcal{L}(s),\mathcal{X})$ and $(M,s) \models \varphi[x/\psi]$ for all $\psi \in \mathcal{L}_n^{K,X,A}(\mathcal{L}(s))$.

We leave it to the reader to make the obvious changes to the remaining clauses.

If C be a (possibly empty) subset of $\{r,t,e\}$, $\mathcal{N}_n^C(\Phi,\mathcal{X})$ be the set of all extended awareness structures such that awareness satisfies *agpp* and *ka* and the possibility correspondence is reflexive, transitive, and Euclidean if these properties are in C. We say that a formula φ is *valid in a class $\mathcal{N}$ of extended awareness structures* if, for all extended awareness structures $M \in \mathcal{N}$ and worlds s such that $\Phi(\varphi) \subseteq \mathcal{L}(s)$, $(M,s) \models \varphi$. (This is essentially the notion of weak validity defined in [Halpern and Rêgo 2008].)

4 AXIOMATIZATION

In this section, we provide a sound and complete axiomatization of the logics described in the previous section. It turns out to be easier to start with the language $\mathcal{L}_n^{\forall,X,A}(\Phi,\mathcal{X})$. All the axioms and inference rules of $\mathrm{AX}^{\mathrm{X,A},\forall}$ continue to be sound in extended awareness structures, except for Barcan_X and FA_X. In a world s where $\mathcal{L}(s) = p$ and agent 1 is aware of p, it is easy to see that $\forall x X_i A_i x$ holds. But if agent 1 considers possible a world t such that $\mathcal{L}(t) = \{p,q\}$, it is easy to see that $X_i \forall x A_i x$ does not hold at s. Similarly, if in world t, agent 1 considers s possible, then $\neg\forall x A_i x$ holds at t, but $X_i \neg\forall x A_i x$ does not. Thus, Barcan_X does not hold at s, and FA_X does not hold at t. We instead use the following variants of Barcan_X and FA_X, which are sound in this framework:

Barcan^*_X. $(A_i(\forall x\varphi) \wedge \forall x(A_i x \Rightarrow X_i\varphi)) \Rightarrow X_i(\forall x A_i x \Rightarrow \forall x\varphi)$.

FA^*_X. $\forall x \neg A_i x \Rightarrow X_i \forall x \neg A_i x$.

Let $\mathrm{AX}_{\mathrm{e}}^{\mathrm{X,A},\forall}$ be the result of replacing FA_X and Barcan_X in $\mathrm{AX}^{\mathrm{X,A},\forall}$ by FA^*_X and Barcan^*_X (the e here stands for "extended").

Theorem 4.1: *If $\mathcal{C}_X$ is a (possibly empty) subset of $\{\mathrm{T_X}, \mathrm{4_X}, \mathrm{5_X}\}$ and C is the corresponding subset of $\{r,t,e\}$, then $\mathrm{AX}_{\mathrm{e}}^{\mathrm{X,A},\forall} \cup \mathcal{C}_\mathrm{X}$ is a sound and complete axiomatization of the language $\mathcal{L}_n^{\forall,X,A}(\Phi,\mathcal{X})$ with respect to $\mathcal{N}_n^C(\Phi,\mathcal{X})$.*

The completeness proof is similar in spirit to that of HR, with some additional complications arising from the interaction between quantification and the fact that different languages are associated with different worlds. What is surprisingly difficult in this case is soundness, specifically, for MP. For suppose that M is a structure in $\mathcal{N}_n(\Phi,\mathcal{X})$ such that neither $\neg\varphi$ nor $\neg(\varphi \Rightarrow \psi)$ are true at any world in M. We want to show that $\neg\psi$ is not true at any world in M. This is easy to show if $\Phi(\psi) \subset \Phi(\varphi)$. For if s is a world such that $\Phi(\psi) \subseteq \mathcal{L}(s)$, it must be the case that both φ and $\varphi \Rightarrow \psi$ are true at s, and hence so is ψ. However, if φ has some primitive propositions that are not in ψ, it is a priori possible that $\neg\psi$ holds at a world where neither φ nor $\varphi \Rightarrow \psi$ is defined. Indeed, this can happen if Φ is finite. For example, if $\Phi = \{p,q\}$, then it is easy to construct a structure $M \in \mathcal{N}_n(\Phi,X)$ where both $A_i p \wedge A_i q$ and $(A_i p \wedge A_i q) \Rightarrow \forall x A_i x$ are never false, but $\forall x A_i x$ is false at some world in M. As we show, this cannot happen if Φ is infinite. This in turn involves proving a general substitution property: if φ is valid and ψ is a quantifier-free sentence, then $\varphi[q/\psi]$ is valid. (We remark that the substitution property also fails if Φ is finite.) See the full paper for details. Proofs for all other results stated in this abstract can also be found in the full paper.

Using different languages has a greater impact on the axioms for K_i than it does for X_i. For example, as we would expect, Barcan does not hold, for essentially the same reason that Barcan_X does not hold. More interestingly, NKA, 5, and Gen_K do not hold either. For example, if $\neg K_i p$ is true at a world s because $p \notin \mathcal{L}(t)$ for some world t that i considers possible at s, then $K_i \neg K_i p$ will not hold at s, even if the $\mathcal{K}_i$ relation is an equivalence relation. Indeed, the properties of K_i in this framework become quite close to the properties of the explicit knowledge operator X_i in the original FH framework, provided we define the appropriate variant of awareness.

Let $A_i^*(\varphi)$ be an abbreviation for the formula $K_i(\varphi \vee \neg\varphi)$. Intuitively, the formula $A_i^*(\varphi)$ captures the property that φ is defined at all worlds considered possible by agent i. Let AGPP*, XA*, A0*, 5*, Barcan*, FA*, and Gen* be the result of replacing X_i by K_i and A_i by A_i^* in AGPP, XA, $A0_X$, 5_X, Barcan_X^*, FA_X^*, and Gen_X, respectively. It is easy to see that AGPP*, A0*, and Gen* are valid in extended awareness structures; XA*, 5*, Barcan*, and FA* are not. For example, suppose that p is defined in all worlds that agent i considers possible at s, so that $A_i^* p$ holds at s. If there is some world t that agent i considers possible at s and a world u that agent i considers possible at t where p is not defined, then $A_i^* p$ does not hold at t, so $K_i A_i^* p$ does not hold at s. It is easy to show that XA* holds if the $\mathcal{K}_i$ relation is transitive. Similar arguments show that 5*, Barcan*, and FA* do not hold in general, but are valid if $\mathcal{K}_i$ is Euclidean and (in the case of Barcan* and FA*) reflexive. We summarize these observations in the following proposition:

Proposition 4.2:

(a) XA is valid in* $\mathcal{N}_n^t(\Phi, \mathcal{X})$.

(b) Barcan is valid in* $\mathcal{N}_n^{r,e}(\Phi, \mathcal{X})$.

(c) FA is valid in* $\mathcal{N}_n^{r,e}(\Phi, \mathcal{X})$.

(d) 5 is valid in* $\mathcal{N}_n^{e}(\Phi, \mathcal{X})$.

In light of Proposition 4.2, for ease of exposition, we restrict attention for the rest of this section to structures in $\mathcal{N}_n^{r,t,e}(\Phi, \mathcal{X})$. Assuming that the possibility relation is an equivalence relation is standard when modeling knowledge in any case. Let $\mathrm{AX}_e^{K,X,A,A^*,\forall}$ be the result of replacing Gen_K and Barcan in $\mathrm{AX}^{K,X,A,\forall}$ by Gen* and Barcan*, respectively, and adding the axioms AGPP*, A0*, and FA* for reasoning about A_i^*. (We do not need the axiom XA*; it follows from 4 in transitive structures.) Let $\mathrm{AX}_e^{K,A^*,\forall}$ consist of the axioms in $\mathrm{AX}_e^{K,X,A,A^*,\forall}$ except for those that mention X_i or A_i; that is, $\mathrm{AX}_e^{K,A^*,\forall} = \mathrm{AX}_e^{K,X,A,A^*,\forall} - \{$AGPP, KA, NKA, A0$\}$. Note that $\mathrm{AX}_e^{K,A^*,\forall}$ is the result of replacing X_i by K_i and A_i by A_i^* in $\mathrm{AX}_e^{X,A,\forall}$ (except that the analogue of XA is not needed). Finally, let AX_e^{K,A^*} consist of the axioms and rules in $\mathrm{AX}_e^{K,A^*,\forall}$ except for the ones that mention quantification; that is, $\mathrm{AX}_e^{K,A^*} = \{$Prop, AGPP*, K, Gen*, A0*$\}$. We use AX_e^{K,A^*} to compare our results to those of HMS.

Theorem 4.3:

(a) $\mathrm{AX}_e^{K,X,A,A^*,\forall} \cup \{T, 4, 5^*\}$ *is a sound and complete axiomatization of the sentences in* $\mathcal{L}_n^{\forall,K,X,A}(\Phi, \mathcal{X})$ *with respect to* $\mathcal{N}_n^{r,e,t}(\Phi, \mathcal{X})$.

(b) $\mathrm{AX}_e^{K,A^*,\forall} \cup \{T, 4, 5^*\}$ *is a sound and complete axiomatization of the sentences in* $\mathcal{L}_n^{\forall,K}(\Phi, \mathcal{X})$ *with respect to* $\mathcal{N}_n^{r,t,e}(\Phi, \mathcal{X})$.

(c) $\mathrm{AX}_e^{K,A^*} \cup \{T, 4, 5^*\}$ *is a sound and complete axiomatization of* $\mathcal{L}_n^{K}(\Phi)$ *with respect to* $\mathcal{N}_n^{r,t,e}(\Phi)$.

Since, as we observed above, $\mathrm{AX}_e^{K,A^*,\forall}$ is essentially the result of replacing X_i by K_i and A_i by A_i^* in $\mathrm{AX}_e^{X,A,\forall}$, Theorem 4.3(b) makes precise the sense in which K_i acts like X_i with respect to A_i^*.

5 DISCUSSION

Just as in our framework, in the HMS and MR approach, a (propositional) language is associated with each world. However, HMS and MR define awareness of φ as an abbreviation of $K_i\varphi \vee K_i\neg K_i\varphi$. In order to compare our approach to that of HMS and MR, we first compare the definitions of awareness. Let $A_i'\varphi$ be an abbreviation for the formula $K_i\varphi \vee K_i\neg K_i\varphi$. The following result says that for extended awareness structures that are Euclidean, $A_i^*\varphi$ is equivalent to $A_i'\varphi$.

Proposition 5.1 : *If* $M = (S, \mathcal{L}, \pi, \mathcal{K}_1, ..., \mathcal{K}_n, \mathcal{A}_1, \ldots, \mathcal{A}_n)$ *is a Euclidean extended awareness structure, then for all* $s \in S$ *and all sentences* $\varphi \in \mathcal{L}_n^{\forall,K,X,A}(\Phi, \mathcal{X})$,

$$(M, s) \models A_i^*\varphi \Leftrightarrow A_i'\varphi.$$

Proof: Suppose that $(M, s) \models K_i(\varphi \vee \neg\varphi) \wedge \neg K_i\varphi$. It follows that $\Phi(\varphi) \subseteq \mathcal{L}(s)$, $\Phi(\varphi) \subseteq \mathcal{L}(t)$ for all t such that $(s, t) \in \mathcal{K}_i$, and that there exists a world t such that $(s, t) \in \mathcal{K}_i$ and $(M, t) \models \neg\varphi$. Let u be an arbitrary world such that $(s, u) \in \mathcal{K}_i$. Since $\mathcal{K}_i$ is Euclidean, it follows that $(u, t) \in \mathcal{K}_i$. Thus, $(M, u) \models \neg K_i\varphi$, so $(M, s) \models K_i\neg K_i\varphi$. It follows that $(M, s) \models A_i'\varphi$, as desired.

For the converse, suppose that $(M, s) \models A_i'\varphi$. If either $(M, s) \models K_i\varphi$ or $(M, s) \models K_i\neg K_i\varphi$, then $\Phi(\varphi) \subseteq \mathcal{L}(s)$, and if $(s, t) \in \mathcal{K}_i$, we have that $\Phi(\varphi) \subseteq \mathcal{L}(t)$. Therefore, $(M, s) \models A_i^*\varphi$. ∎

In [Halpern and Rêgo 2008], we showed that $\mathrm{AX}_e^{K,A^*} \cup \{T, 4, 5^*\}$ provides a sound and complete axiomatization of the structures used by HMS where the possibility relations are Euclidean, transitive, and reflexive, with one difference: A'_i is used for awareness instead of A^*_i. However, by Proposition 5.1, in $\mathcal{N}^e_n$, A^*_i and A'_i are equivalent. Thus, for the class of structures of most interest, we are able to get all the properties of the HMS approach; moreover, we can extend to allow for reasoning about knowledge of unawareness. It is not clear how to capture knowledge of unawareness directly in the HMS approach.

It remains to consider the relationship between A_i and A^*_i. Let $\mathcal{A}^*_i(s)$ be the set of sentences that are defined at all worlds considered possible by agent i in world s; that is, $\varphi \in \mathcal{A}^*_i(s)$ iff $(M,s) \models A^*_i\varphi$. Assuming that agents know what they are aware of, we have that if $(s,t) \in \mathcal{K}_i$, then $\mathcal{A}_i(s) = \mathcal{A}_i(t)$. Thus, it follows that $\mathcal{A}_i(s) \subseteq \mathcal{A}^*_i(s)$. For if $\varphi \in \mathcal{A}_i(s)$, then $\Phi(\varphi) \subseteq \mathcal{L}(t)$ for all t such that $(s,t) \in \mathcal{K}_i$, so $(M,s) \models A^*_i(\varphi)$.

We get the opposite inclusion by assuming the following natural connection between an agent's awareness function and the language in the worlds that he considers possible:

- **LA:** If $p \notin \mathcal{A}_i(s)$, then $p \notin \mathcal{L}(t)$ for some t such that $(s,t) \in \mathcal{K}_i$.

It is immediate that in models that satisfy **LA** (and *agpp*), $\mathcal{A}_i(s) \supseteq \mathcal{A}^*_i(s)$ for all agents i and worlds s. Thus, under minimal assumptions, $\mathcal{A}^*_i(s) = \mathcal{A}_i(s)$.

The bottom line here is that under the standard assumptions in the economics literature, together with the minimal assumption **LA**, all the notions of awareness coincide. We do not need to consider a syntactic notion of awareness at all. However, as pointed out by FH, there are other notions of awareness that may be relevant; in particular, a more computational notion of awareness is of interest. For such a notion, an axiom such as AGPP does not seem appropriate. We leave the problem of finding axioms that characterize a more computational notion of awareness in this framework to future work.

We conclude with some comments on awareness and language. If we think of propositions $p \in \mathcal{L}(t) - \mathcal{A}_i(s)$ as just being labels or names for concepts that agent i is not aware of but i understands other agents might be aware of, **LA** is just saying that i should not use the same label in all worlds that he considers possible. It is important that an agent can use different labels for formulas that he is unaware of. A world where an agent is unaware of two primitive propositions is different from a world where an agent is unaware of only one primitive proposition. For example, to express the fact that in world s agent agent 1 considers it possible that (1) there is a formula that he is unaware that agent 2 is aware of and (2) there is a formula that both he and agent 2 are unaware of that agent 3 is aware of, agent 1 needs to consider possible a world t with at least two primitive propositions in $\mathcal{L}(t) - \mathcal{A}_1(s)$. Needless to say, reasoning about such lack of awareness might be critical in a decision-theoretic context.

The fact that the primitive propositions that an agent is not aware of are simply labels means that switching the labels does not affect what the agent knows or believes. More precisely, given a model $M = (S, \mathcal{L}, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{A}_1, \ldots, \mathcal{A}_n, \pi)$, let M' be identical to M except that the roles of the primitive propositions p and p' are interchanged. More formally, $M' = (S, \mathcal{L}', \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{A}'_1, \ldots, \mathcal{A}'_n, \pi')$, where, for all worlds $s \in S$, we have

- $\mathcal{L}(s) - \{p, p'\} = \mathcal{L}'(s) - \{p, p'\}$;
- $p \in \mathcal{L}'(s)$ iff $p' \in \mathcal{L}(s)$, and $p' \in \mathcal{L}'(s)$ iff $p \in \mathcal{L}(s)$;
- $\pi(s,q) = \pi'(s,q)$ for all $q \in \mathcal{L}(s) - \{p, p'\}$;
- if $p \in \mathcal{L}(s)$, then $\pi(s,p) = \pi'(s,p')$, and if $p' \in \mathcal{L}(s)$, then $\pi(s,p') = \pi'(s,p)$;
- if φ is a formula that mentions neither p nor p', then $\varphi \in \mathcal{A}_i(s)$ iff $\varphi \in \mathcal{A}'_i(s)$;
- for any formula φ that mentions either p or p', $\varphi \in \mathcal{A}_i(s)$ iff $\varphi[p \leftrightarrow p'] \in \mathcal{A}'_i(s)$, where $\varphi[p \leftrightarrow p']$ is the result of replacing all occurrences of p in φ by p' and all occurrences of p' by p.

It is easy to see that for all worlds s, $(M,s) \models \varphi$ iff $(M',s) \models \varphi[p \leftrightarrow p']$. In particular, this means that if neither p nor p' is in $\mathcal{L}(s)$, then for all formulas, $(M,s) \models \varphi$ iff $(M',s) \models \varphi$. Thus, switching labels of propositions that are not in $\mathcal{L}(s)$ has no impact on what is true at s.

We remark that the use of labels here is similar in spirit to our use of *virtual moves* in [Halpern and Rêgo 2006a] to model moves that a player is aware that he is unaware of.

Although switching labels of propositions that are not in $\mathcal{L}(s)$ has no impact on what is true at s, changing the truth value of a primitive proposition that an agent is not aware at s may have some impact on what the agent explicitly knows at s. Note that we allow agents to have some partial information about formulas that they are unaware of. We certainly want to allow agent 1 to know that there is a formula that agent 2 is aware of that he (agent 1) is unaware of; indeed, capturing a situation like this was one of our primary motivations for introducing knowledge of lack of awareness. But we also want to allow agent 1 to know that agent 2 is not only aware of the formula, but knows that it is true; that is, we want $X_1(\exists x(\neg A_1(x) \wedge K_2(x)))$ to be consistent. There may come a point when an agent has so much partial information about a formula he is unaware of that, although he cannot talk about it explicitly in his language, he can describe it sufficiently well to communicate about it. When this happens in natural language, people will come up with a name for a concept and add it to their language. We have not addressed the dynamics of language change here, but we believe that this is a topic that deserves further research.

Acknowledgments The first author is supported in part by NSF grants ITR-0325453, IIS-0534064, and IIS-0812045, and by AFOSR grants FA9550-08-1-0438 and FA9550-05-1-0055. The second author is supported in part by FACEPE under grants APQ-0150-1.02/06 and APQ-0219-3.08/08, and by MCT/CNPq under grant 475634/2007-1.

References

Board, O. and K.-S. Chung (2009). Object-based unawareness: theory and applications. unpublished manuscript.

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

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

Halpern, J. Y. and L. C. Rêgo (2006a). Extensive games with possibly unaware players. In *Proc. Fifth International Joint Conference on Autonomous Agents and Multiagent Systems*, pp. 744–751. Full version available at arxiv.org/abs/0704.2014.

Halpern, J. Y. and L. C. Rêgo (2006b). Reasoning about knowledge of unawareness. In *Principles of Knowledge Representation and Reasoning: Proc. Tenth International Conference (KR '06)*, pp. 6–13. To appear, *Games and Economic Behavior*.

Halpern, J. Y. and L. C. Rêgo (2008). Interactive unawareness revisited. *Games and Economic Behavior 62*(1), 232–262.

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

Heifetz, A., M. Meier, and B. Schipper (2008). A canonical model for interactive unawareness. *Games and Economic Behavior 62*, 305–324.

Hintikka, J. (1962). *Knowledge and Belief*. Ithaca, N.Y.: Cornell University Press.

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

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

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