Centre for Climate Change Economics and Policy

Grantham Research Institute on Climate Change and the Environment

Unawareness with 'possible' possible worlds

Oliver Walker

December 2011

Centre for Climate Change Economics and Policy
Working Paper No. 78

Grantham Research Institute on Climate Change and the Environment
Working Paper No. 69

The Grantham Foundation

ESRC Economic & Social Research Council

University of Leeds

LSE The London School of Economics and Political Science

The Centre for Climate Change Economics and Policy (CCCEP) was established by the University of Leeds and the London School of Economics and Political Science in 2008 to advance public and private action on climate change through innovative, rigorous research. The Centre is funded by the UK Economic and Social Research Council and has five inter-linked research programmes:

1. Developing climate science and economics
2. Climate change governance for a new global deal
3. Adaptation to climate change and human development
4. Governments, markets and climate change mitigation
5. The Munich Re Programme - Evaluating the economics of climate risks and opportunities in the insurance sector

More information about the Centre for Climate Change Economics and Policy can be found at: http://www.cccep.ac.uk.

The Grantham Research Institute on Climate Change and the Environment was established by the London School of Economics and Political Science in 2008 to bring together international expertise on economics, finance, geography, the environment, international development and political economy to create a world-leading centre for policy-relevant research and training in climate change and the environment. The Institute is funded by the Grantham Foundation for the Protection of the Environment, and has five research programmes:

1. Use of climate science in decision-making
2. Mitigation of climate change (including the roles of carbon markets and low-carbon technologies)
3. Impacts of, and adaptation to, climate change, and its effects on development
4. Governance of climate change
5. Management of forests and ecosystems

More information about the Grantham Research Institute on Climate Change and the Environment can be found at: http://www.lse.ac.uk/grantham.

Unawareness with "Possible" Possible Worlds

Oliver Walker
The Grantham Research Institute on Climate Change and the Environment
London School of Economics
o.j.walker@lse.ac.uk

December 3, 2011

Abstract

Logical structures for modeling agents' reasoning about unawareness are presented where it can hold simultaneously that: (i) agents' beliefs about whether they are fully aware need not be veracious with partitional information; and (ii) the agent is fully aware if and only if she is aware of a fixed domain of formulae. In light of (ii), all states are deemed "possible". Semantics operate in two stages, with belief in the second stage determined by truth in the first stage. Characterization theorems show that, without the first stage, the structures validate the same conditions as those of Halpern and Rego (2009a).

1 Introduction

Being unaware of something is the same as being unable to conceive of it. Unawareness of the sentence ϕ thus implies not only the absence of belief in ϕ but also the absence of belief in *any* sentence involving ϕ, including tautologies such as "ϕ or not ϕ".

Allowing for unawareness in economic models of decision makes it possible to explain an array of phenomena that standard frameworks preclude. For example, suppose Donald is unaware of the possibility of a bank run and chooses to invest his money in shares in HSBC. If one were to give Donald the message, "there will or will not be a bank run on HSBC", standard Bayesian models of decision would deny any possibility of Donald changing his mind as the message is not "informative" in the sense of ruling out any state of affairs. However, given Donald's prior unawareness, the message *does* impact upon Donald's information (put another way, it changes his beliefs) and it therefore seems reasonable to allow Donald to revise his investment strategy in light of receiving it. The question of how unawareness may affect dynamic decision in this way has been studied by Li (2008), Grant and Quiggin (2009), and Karni and Viero (2009).

Donald's unawareness of the possibility of a bank run is an example of what this paper refers to as *specific unawareness*: unawareness of a particular statement or event. Though no agent who was (specifically) unaware of ϕ could ever believe she was unaware of ϕ – to do so would imply initial awareness of ϕ – such an agent could hold beliefs about the more general possibility that there was *something* (non-specific) of which she was unaware. In this paper the phenomenon of holding such beliefs is called *conscious unawareness*.

Conscious unawareness may have a range of interesting economic consequences. Tirole (2009) argues that if contracting parties believe they are unaware, it may be efficient for them to allow scope for renegotiation in case they become aware of more relevant facts in future (that is, it may be efficient to write incomplete contracts). Other applications are interactive: for example, Halpern and Rego (2009a) suggest that a local doctor may refer her patient to a specialist because she believes the specialist is aware of relevant things that she is not. Walker and Dietz (2011) claim that even in static single-agent settings, a decision maker's conscious unawareness may lead her to violate subjective expected utility.

Frameworks for modeling conscious unawareness should be consistent with agents satisfying reasonable *introspective properties*. For instance, as Dekel et al. (1998) argue, it should at least be possible that an agent who is unaware of ϕ does not believe she is unaware of ϕ. Further introspective properties are highlighted by the examples below:

Example 1: Martha is an inexperienced gambler playing a card game with her friends. On the basis of her understanding of the rules of the game, she believes that her current hand cannot be beaten. This belief (and the conception of the rules on which it is founded) is correct. However, when the time comes for her to place a bet, she decides not to gamble, reasoning that there may additional rules of which she is unaware that could cause her to lose money.

Example 2: Eunice is a social worker visiting a family she has known for several years. She interviews the family members and observes their behavior, watching for signs of anything that would be of concern to social services, and then writes a report stating her conclusions. Given her longstanding acquaintance with the family and years of professional experience, she is confident that she is aware of all of the relevant concerns and, furthermore, she has sufficient evidence to conclude that none of them apply to the family. She therefore writes that there is nothing to be worried about. This assessment later proves wrong.

While it is true that both Martha's and Eunice's choices turn out to be contrary to their interests, neither of the examples is intended to portray unreasonable or irrational beliefs or behavior.

Consider Martha's case and let r be one of the rules of the game. It is natural to suppose that the fact of Martha's awareness of r is self-evident to her and thus that she believes she is aware of r. After all, the very fact that she uses r to deduce that she has a winning hand should be enough to convince her that she is aware of it. Given this, an economist modeling a consciously unaware agent might wish to require that for any statement ϕ:

$$\text{The agent is aware of } \phi \implies \text{She believes she is aware of } \phi \tag{1}$$

By contrast, there seems to be little justification for assuming that the *extent* of Martha's awareness is self-evident to her. For although she can have no idea what an additional rule to the game might involve, this could also be the case if there was a rule she was unaware of. Thus, her inexperience alone gives her good grounds to suspect that there she is not fully aware. Models of conscious unawareness should therefore be consistent with:

$$\text{The agent is fully aware} \quad \not\Longrightarrow \quad \text{She believes she is fully aware} \tag{2}$$

Turning to the case of Eunice, similar reasoning applies. She should not be expected to recognize the existence of problems outside her conception since, by their nature, these problems could not be self-evident. And given her close knowledge of the family, it is reasonable for her to suppose that there may be no problems of this kind. This suggests that models of conscious unawareness should satisfy:

$$\text{The agent is not fully aware} \quad \not\Longrightarrow \quad \text{She believes she is not fully aware} \tag{3}$$

However, it turns out that (1) is inconsistent with (2) and (3) in the most straightforward types of model for conscious unawareness. To see this consider the case of Martha again. According to (1) whatever she is aware of, she believes she is aware of, and whatever she believes must be true in every state of the world she considers possible. This means that in every state she regards as possible, her level of awareness is at least as great as it is in the true state of the world. But this implies that she is fully aware in every state she considers possible, which means that she believes she is fully aware, a violation of (2).

This difficulty, which is present in Halpern and Rego (2009a), has been recognized in the literature and several authors[1] have proposed alternative approaches that circumvent the problem. These works resolve the inconsistency by allowing the set of things the agent *could* be aware of to vary from state to state – thus, though Martha is aware of the same things in all states she regards as possible, it need not be that she is aware of everything there is to be aware of in all these states so it does not follow that she believes she is fully aware.

Varying the domain of "things to be aware of" across states of the world in this way allows one to reconcile (1) with (2) and (3), but the resulting state space is no longer of the kind familiar from Savage (1954). In the latter type of state space, states stand for a complete and consistent description of reality, where notions of "completeness" and "consistency" are fixed across states. Yet if the set of "things to be aware of" differs from state to state, it must be that the consistency of the sentence "the agent is fully aware" with a given level of awareness is not fixed. Thus, it cannot be that all states use a single, "objective", criterion of consistency: those that depart from this – for instance, any state used to model Example 2 where the set of rules is greater than it really is – are described in this paper as "impossible" states of the world.

There are at least three reasons why it is desirable to model conscious unawareness using only "possible" states. The first is decision theoretic. In order to characterize

[1] These include Sillari (2008), Board and Chung (2007), and Halpern and Rego (2009b).

a decision maker's beliefs in revealed preference terms[2], the usual approach is to specify payoffs to different courses of action under every state and endow the decision maker with conditional preferences in every state of the world[3]. But if a state of the world is impossible, it is difficult to see how payoffs and conditional preferences under this state could ever be defined.

Second, working with a framework that includes both possible and impossible states of the world complicates any reduction from a linguistic to a set-theoretic representation of the agent's reasoning. As was shown by Dekel et al. (1998), set-theoretic models for non-trivial specific unawareness must enrich the standard Savage framework by describing both what the agent's level of awareness is and what events she regards as possible[4]. Set-theoretic models for conscious unawareness that use impossible states (for example, Board and Chung (2009)) need to elaborate this framework further by specifying, for each state, what degree of awareness amounts to "full" awareness.

Finally, a pragmatic but arguably more important reason for working with only possible states is that this has been the approach used in the overwhelming majority of existing economic theory. If the goal is to integrate conscious unawareness with this body of work in a manner that is accessible to non-experts, it seems sensible to avoid departures from its existing standards wherever this is feasible.

This paper presents a logical structure for modeling conscious unawareness where all states are possible and where (1), (2), and (3) are nonetheless mutually consistent. It achieves this by making the agent's beliefs about whether or not she is fully aware in any state dependent on her *conjecture* about her level of awareness in that state rather than on a comparison between her actual awareness and the domain of "things to be aware of". The semantics of these structures works in two stages. At the first stage, "conjectured truth" is assigned to logical formulae in all states. For any formula that does not include the string "the agent is fully aware", this conjecture takes the same value as it would under conventional semantic rules, but the truth of "the agent is fully aware" may be determined freely. The second stage then assigns truth to formulae in the same way as in Fagin and Halpern (1988), with the exception that an agent believes a formula if and only if she is aware of it and it is conjectured true in every state she considers possible. One way of interpreting this two-stage approach is as allowing the agent's beliefs about "the agent is fully aware" to be determined by some process of *inductive* reasoning (as in Grant and Quiggin (2009)), whereas her other beliefs are arrived at by a process of deduction from what is true (and she is aware of) in the states she deems possible.

As well as relying only on possible states, the framework proposed here has the advantage of simplicity. Whereas other logical structures for modeling conscious unawareness use quantification over propositional variables to assign truth to the

[2]Morris (1996) shows how this may be done in a framework without unawareness. Schipper (2010) and Li (2008) offer revealed-preference characterizations of belief in the presence of specific unawareness. Walker (2011a) does this in a setting that allows for both specific and conscious unawareness.

[3]Though note the contribution of Lipman (2003).

[4]Heifetz et al. (2006) and Li (2009) offer different approaches to this.

formula "the agent is fully aware", in what follows the conjectured truth of this formula is determined much like that of a primitive proposition in standard logic, while its second-stage truth follows from some straightforward consistency conditions. As a result, the logical structures presented below do not need to use explicit quantification.

The remainder of the paper is organized as follows. Section 2 presents the elements of the logical structures including the two-stage semantics, while Section 3 discusses various introspective axioms from elsewhere in the literature. Section 4 characterizes the structures axiomatically under a number of well known restrictions, showing that when the agent's conjecture about "the agent is fully aware" is constrained to be veracious, the structures are equivalent to those in Halpern and Rego (2009a). Section 5 concludes with a brief review of some of the other related literature.

2 Two-Stage Semantics for Awareness and Belief

Consider a formal language, denoted $\mathcal{L}$, for expressing the kind of reasoning that was discussed in the Introduction. $\mathcal{L}$ consists of a countably infinite set of *primitive propositions*, P, the string $\forall xAx$, the logical connectives $\neg$ and $\vee$, the brackets [and], and the modal operators A, L, and B. The primitive propositions are a set of claims about the world that do not concern the decision maker's beliefs or awareness, such as "it will rain tomorrow" or "demand for bread will be high", and the string $\forall xAx$ stands for the claim "the agent is fully aware".

Any member of P and $\forall xAx$ is a *formula* of the language, the full set of which is denoted Φ. Φ is defined formally as the closure of P and $\forall xAx$ under the following formation rules[5]:

$$\begin{aligned} \phi \in \Phi &\implies \neg\phi, A\phi, L\phi, B\phi \in \Phi \\ \phi, \psi \in \Phi &\implies [\phi \vee \psi] \end{aligned}$$

The connectives $\neg$ and $\vee$ respectively stand for "not" and "or", and allow formulae expressing negation and disjunction to be constructed from the members of Φ. Thus, if ϕ and ψ are formulae, then so are $\neg\phi$ and $[\phi \vee \psi]$. The modal operators A, L, and B, read "the agent is aware of the formula", "the agent implicitly believes", and "the agent (explicitly) believes", allow one to construct statements about the agent's epistemic status (how such statements should be interpreted is discussed below). As is standard: $[\phi \wedge \psi]$ ("ϕ and ψ"), $[\phi \Rightarrow \psi]$ ("ϕ implies ψ"), and $[\phi \Leftrightarrow \psi]$ ("ϕ if and only if ψ") will be used as shorthand for $\neg[\neg\phi \vee \neg\psi]$, $[\neg\phi \vee \psi]$ and $[[\phi \Rightarrow \psi] \wedge [\psi \Rightarrow \phi]]$ respectively throughout this paper, and brackets will generally be suppressed unless doing so leads to ambiguity.

A *structure*, D, in $\mathcal{L}$, the full set of which is denoted $\mathcal{D}$, is a tuple $\{\Omega, \mathcal{V}, \mathcal{P}, \mathcal{A}, \mathcal{X}\}$. Ω is a state space consisting of possible worlds or states of affairs. $\mathcal{V} : \Omega \to 2^P$ is a

[5]It should be noted that $\mathcal{L}$, unlike its counterparts in other papers in this literature, is not a proper first order language in that it does not contain propositional variables and does not allow quantification over any formula. Since the only quantified formulae of interest in this setting is $\forall xAx$, this additional expressive power would be redundant here.

truth assignment to the primitive propositions, listing those that are true in each state, and $\mathcal{P} : \Omega \to 2^\Omega$ is a *possibility correspondence*, where $\mathcal{P}(\omega)$ stands for the set of states that the agent considers possible when the true state is ω. $\mathcal{A} : \Omega \to 2^P$ is an *awareness correspondence*, mapping each state to the set of propositions that the agent is aware of in that state and, finally, $\mathcal{X} : \Omega \to \{T, F\}$ describes, for each state, the *conjecture* the agent has about whether he is fully aware in each state. If $\mathcal{X}(\omega) = T$ then the agent imagines that in state ω his awareness is exhaustive (i.e. his impression of $\forall x A x$ is that it is true in ω), while $\mathcal{X}(\omega) = F$ means that he conjectures that $\forall x A x$ is false in ω.

In what follows, a *subformula* of any formula ϕ is any consecutive string of characters in ϕ that is itself a formula of $\mathcal{L}$. For example, if $\phi = A[p \vee Bq]$, then $A[p \vee Bq]$, $p \vee Bq$, Bq, p, and q are the subformulae of ϕ. Use $\mathrm{Sub}(\phi)$ for the set of subformulae of any $\phi \in \Phi$ and define the *referents* of ϕ as $\mathrm{Ref}(\phi) := \mathrm{Sub}(\phi) \cap P$ (that is, $\mathrm{Ref}(\phi)$ is the set of primitive propositions that are subformulae of ϕ).

The *semantics* of $\mathcal{D}$ comprises a set of rules that assign truth to every formula in Φ in every state in Ω according to V, $\mathcal{P}$, $\mathcal{A}$, and $\mathcal{X}$. The rules presented here are unusual in that they operate in two stages: first, there is a set that establishes what is "conjectured true" or "stage-one true" in each state; and second, there are rules for assigning "stage-two truth", or simply "truth" to the formulae. The role of sage-one truth is to determine what the agent believes and at this level the truth of the formula $\forall x A x$ depends on the agent's conjecture rather than the extent of his awareness. Stage-two truth differs in that $\forall x A x$ is true at this level if and only if the agent is in fact aware of all propositions. As will be shown, this divergence between the "conjectured" way in which stage-one truth is assigned to $\forall x A x$ and the "objective" standard applied to it for stage-two truth is what allows the members of $\mathcal{D}$ to satisfy the various introspection properties advocated in the Introduction.

Use the operator $V_1 : \Omega \times \Phi \to \{0, 1\}$ to denote what formulae are stage-one true at each state of a given structure. $V(\omega, \phi) = 1$ is read "ϕ is stage-one true in state ω" and $V(\omega, \phi) = 0$ means ϕ is first-stage false at ω. The operator is defined inductively as follows:

- S1. $V_1(\omega, p) = 1$ iff $p \in \mathcal{V}(\omega)$ where $p \in P$;
- S2. $V_1(\omega, \forall x A x) = 1$ iff $\mathcal{X}(\omega) = T$;
- S3. $V_1(\omega, \neg\phi) = 1 - V_1(\omega, \phi)$;
- S4. $V_1(\omega, [\phi \vee \psi]) = \max\{V_1(\omega, \phi), V_1(\omega, \psi)\}$;
- S5. $V_1(\omega, A\phi) = 1$ iff $\mathrm{Ref}(\phi) \subseteq \mathcal{A}(\omega)$;
- S6. $V_1(\omega, L\phi) = 1$ iff $V_1(\omega', \phi) = 1$ for all $\omega' \in \mathcal{P}(\omega)$; and
- S7. $V_1(\omega, B\phi) = 1$ iff $V_1(\omega, A\phi) = 1$ and $V_1(\omega, L\phi) = 1$.

S1, S3, and S4 are familiar from propositional calculus. S1 requires that p is stage-one true in state ω if and only if it is true in that state according to $\mathcal{V}$, while S3 ensures the negation of ϕ is stage-one true iff ϕ is stage-one false and S4 makes $[\phi \vee \psi]$ stage-one true iff at least one of its disjuncts is stage-one true.

S5-S7 govern awareness and belief and operate in an analogous fashion to their equivalents in Fagin and Halpern (1988) and Halpern (2001). S5 implies that it is stage-one true that agent is aware of ϕ in ω iff he is aware of all of ϕ's referents in ω. This is what Halpern (2001) calls "awareness generated by primitive propositions". S6 says that the agent implicitly believes ϕ in ω iff ϕ is stage-one true in every state he considers possible at ω, and S7 then identifies belief with implicit belief plus awareness. S5-S7 suggest that implicit belief can be interpreted as a component of belief rather than an epistemic property in its own right: part of believing ϕ is considering possible only states where ϕ is true, but there is no state of mind that corresponds this.

The novel part of the stage-one semantics - and the respect in which these semantics are "conjectured" - is S2. This rule states that $\forall x Ax$ is stage-one true in ω if and only if the agent's conjecture is that his awareness is comprehensive in ω. The way in which the stage-one truth of $\forall x Ax$ (and other formulae of which it is a subformula) in any given state is determined therefore differs from other formulae in that it depends on the agent's conjecture rather than features of the state itself.

Second-stage truth, which determines in which states any given formula does hold true, is assigned according to a second set of rules. Where $V(\omega, \phi) = 1$ means ϕ is stage-two true (or simply "true") in state ω and $V(\omega, \phi) = 0$ means ϕ is false at ω, these rules are as follows:

O1. $V(\omega, p) = 1$ iff $p \in \mathcal{V}(\omega)$ where $p \in P$;

O2. $V(\omega, \forall x Ax) = 1$ iff $\mathcal{A}(\omega) = P$;

O3. $V(\omega, \neg\phi) = 1 - V(\omega, \phi)$;

O4. $V(\omega, [\phi \vee \psi]) = \max\{V(\omega, \phi), V(\omega, \psi)\}$;

O5. $V(\omega, A\phi) = 1$ iff $\mathrm{Ref}(\phi) \subseteq \mathcal{A}(\omega)$;

O6. $V(\omega, L\phi) = 1$ iff $V_1(\omega', \phi) = 1$ for all $\omega' \in \mathcal{P}(\omega)$; and

O7. $V(\omega, B\phi) = 1$ iff $V(\omega, A\phi) = 1$ and $V(\omega, L\phi) = 1$.

Most of the rules function in essentially the same manner as their equivalents in S1-S7. The exceptions are O2 and O7. O2 uses an "objective" criterion in place of the agent's conjecture to assign truth to the formula $\forall x Ax$, stating simply that $\forall x Ax$ is true in ω if and only if the agent is aware of all propositions (and therefore, by O6, all formulae) in ω. O7 marks the point of contact between the first- and second-stage semantics, and implies that the agent implicitly believes ϕ in ω if and only if ϕ is stage-one true when in every state he considers possible at ω. Thus, whether or not an agent believes a formula depends on its stage-one truth in the states he regards as possible and, for those formulae in which $\forall x Ax$ appears as a subformula, this can be different from its second-stage truth.

Note that all of the states in any structure in $\mathcal{D}$ are "possible" in the sense outlined in the Introduction as the consistency of $\forall x Ax$ with any given level of awareness is the same in all states.

3 Properties of awareness

It remains to be shown what the intrinsic properties of $\mathcal{D}$ are, and whether its structures can be used to describe the sort of reasoning described in the Introduction. A useful concept for investigating these matters is *validity*. A set of formulae Δ will be said to be valid in the structure D iff, for every $\phi \in \Delta$, $V(\omega, \phi) = 1$ for all ω in D's state space. Δ is then valid in $\mathcal{D}^* \subseteq \mathcal{D}$ iff it is valid in all $D \in \mathcal{D}^*$, and if it is valid in $\mathcal{D}$, it will be described simply as being valid.

Many of the sets of formulae described in this section are presented as *schemes*, where a scheme is a formula, certain subformulae of which can be replaced uniformly by any other formula from a particular domain. For example, the scheme $\phi \vee [\psi \wedge \neg\phi]$ for any $\phi, \psi \in \Phi$ contains $p \vee [q \wedge \neg p]$ but not $p \vee [B\forall x Ax \wedge \neg Lr]$, since ϕ is not uniformly replaced by any formula in the latter case. A scheme is valid in some $\mathcal{D}^* \subseteq \mathcal{D}$ whenever every instance of it is valid in $\mathcal{D}^*$.

There are certain schemes whose validity is necessary if the A and B operators are to be plausibly interpreted as representing awareness and belief across all $\mathcal{D}$. Consider the following sets, where ϕ can be any member of Φ:

PL $\neg A\phi \implies \neg B\phi \wedge \neg B\neg B\phi$; and

AU $\neg A\phi \implies \neg A\neg A\phi$.

PL, which stands for "plausibility"[6], is a formal statement of the fact that whenever an agent is unaware of something, he cannot believe it and cannot believe that he does not believe it. Similarly, **AU** (for "awareness-unawareness introspection") expresses the fact that an agent who is unaware of ϕ cannot be aware that he is unaware of ϕ. Any structure in which either **PL** or **AU** were not valid would not be suitable for modelling belief and awareness, but it is easy to show that both are valid in $\mathcal{D}$.

The validity of other sets of formulae shows what sort of awareness the structures in $\mathcal{D}$ can model. Take the following formulae, where once again ϕ and ψ can be any formulae:

A0 $A\forall x Ax$;

A1 $A\neg\phi \iff A\phi$

A2 $A[\phi \vee \psi] \iff A\phi \wedge A\psi$;

A3 $AL\phi \iff A\phi$;

[6]This terminology, and that of AU introspection, is due to Dekel et al. (1998).

A4 $AA\phi \Longleftrightarrow A\phi$;

A5 $AB\phi \Longleftrightarrow A\phi$; and

A6 $\forall x Ax \Longrightarrow Ap$ for all $p \in P$.

A0-A5 jointly correspond to "awareness generated by primitive propositions" - $A\phi$ iff Ap for all $p \in \mathrm{Ref}(\phi)$ - which seems to be a natural form of awareness to attribute to an economic agent[7]. One way of interpreting **A0-A5** is the assumption that the agent is aware of all aspects of the grammar of $\mathcal{L}$ (i.e. the rules for constructing formulae) but not necessarily all of the vocabulary (i.e. the propositions). **A6** then formalizes the interpretation of $\forall x Ax$ as "the agent is aware of all formulae". It is straightforward to show that all of **A0-A6** are valid in $\mathcal{D}$ and that **A1** and **A4** entail **AU**.

A further group of formulae can be regarded as conditions on the rationality of the agent under consideration. Where ϕ can be any member of Φ, none of the following is valid in $\mathcal{D}$:

PI* $B\phi \Longrightarrow BB\phi$;

NI* $\neg B\phi \wedge A\phi \Longrightarrow B\neg B\phi$; and

BA $A\phi \Longrightarrow BA\phi$.

PI* (for "positive introspection") and **NI*** ("negative introspection") respectively state that the agent believes he believes whatever formulae he believes and that he believes he does not believe whatever formulae he does is aware of but does not believe. **BA**, which as (1) was endorsed in the Introduction, then says that he believes he is aware of whatever he is aware of. Considered together, the three properties amount to the agent knowing the content of his own mind: he knows the formulae he is aware of and knows whether or not he believes each of these formulae. If part of being rational is deducing what is self-evident and the content of the agent's mind is self-evident to the him, then **PI**, **NI**, and **BA** function as rationality conditions.

In some economic theories, rationality also commits the agent to having only true beliefs. The argument for this is that in order to hold false beliefs, an agent must misinterpret whatever evidence of the true state of the world he has at his disposal, and such an agent could not be said to deduce beliefs from evidence in a rational fashion. However, since in $\mathcal{D}$ the agent's beliefs concerning the formula $\forall x Ax$ are determined by a conjecture rather than true features of the states he considers possible, this veracity condition is applied only to a restricted domain. Define this domain, denoted $\Phi^{\forall -}$, inductively as follows:

$$\begin{aligned}
p \in P &\quad\Longrightarrow\quad p \in \Phi^{\forall -} \\
\phi \in \{A\psi, L\psi, B\psi\} &\quad\Longrightarrow\quad \phi \in \Phi^{\forall -} \\
\phi \in \tilde{\Phi} &\quad\Longrightarrow\quad \top\phi, \neg\phi \in \Phi^{\forall -} \\
\phi, \chi \in \tilde{\Phi} &\quad\Longrightarrow\quad [\phi \vee \chi] \in \Phi^{\forall -}
\end{aligned}$$

[7] Halpern (2001) describes other types of awareness, which could be accommodated in structures very similar to $\mathcal{D}$.

where ψ may be any formula in Φ. $\Phi^{\forall -}$ is the set of formulae whose stage-two truth in any state in any $D \in \mathcal{D}$ is always the same as their stage-one truth in that state. The rationality condition can now be stated as:

T* $B\phi \Longrightarrow \phi$ for any $\phi \in \Phi^{\forall -}$;

T* is not valid in $\mathcal{D}$.

Finally, there are some formulae whose validity would entail the agent's having overly strong powers of cognizance. Consider the following:

P∀ $\forall x Ax \Longrightarrow B\forall x Ax$; and

N∀ $\neg\forall x Ax \Longrightarrow B\neg\forall x Ax$.

P∀ and **N∀** formalize the notion that an agent should always know whether or not he is aware of everything. As was argued in the Chapter 1, this is not a compelling property in a wide range of economic applications, so structures in which either formula is valid should be avoided when modeling such environments. Neither formula is valid in $\mathcal{D}$.

4 Characterization results

Characterization theorems identify those formulae that are valid in a given class of structures in $\mathcal{D}$. They thus show precisely what can and cannot be modeled using the structures in question. The purpose of this section is to state a characterization theorem for the whole of $\mathcal{D}$, as well as for certain subsets of $\mathcal{D}$ in which the rationality conditions introduced above are validated. A further characterization is provided for the structures in $\mathcal{D}$ that correspond to those in Halpern and Rego (2009a), and it is shown that these cannot be reconciled with the rationality conditions without entailing **P∀** and **N∀**.

The sets of formulae that characterize various subsets of $\mathcal{D}$ will be presented as *axiom systems*. An axiom system consists of a number of foundational schemes known as axioms and a set of deductive rules, which take the form "from $\phi_1, \ldots, \phi_n$ conclude ψ". A *proof* of a formula, ϕ_n, in an axiom system is a finite list of formulae, $\phi_1, \ldots, \phi_n$, such that for each ϕ_i in the list, either ϕ_i is an instance of an axiom scheme, or there is a deductive rule that says ϕ_i can be concluded from some subset of $\phi_1, \ldots, \phi_{i-1}$. A *theorem* of a system is any formula for which there exists a proof within that system, and a system characterizes some $\mathcal{D}^* \subseteq \mathcal{D}$ iff the set of theorems of the system is identical to the set of formulae that are valid in $\mathcal{D}^*$.

All of the axiom systems to be considered in this section include the following:

PC Any tautology in propositional calculus; and

K $[L\phi \wedge L[\phi \Rightarrow \psi]] \Longrightarrow L\psi$ (Distribution) where $\phi, \psi \in \Phi$. The tautologies of propositional calculus are those formulae that are valid in virtue of rules O3 and O4. Distribution states that if an agent implicitly believes an implication and its antecedent, he must implicitly believe its consequent. **PC** and **K** are common to all "normal" axiom systems in modal logic, where belief may be characterized using a possibility correspondence(see Hughes and Cresswell (1996) or Fagin et al. (1995) for more details).

A further axiom that is present in all the systems discussed here is equivalent to the definition of belief given in O7:

A7 $B\phi \Longleftrightarrow A\phi \wedge L\phi$

for any $\phi \in \Phi$. Note that **A7** in combination with **A2** and **A6** entails **PL**.

All axiom systems in this paper include the following set of deductive rules:

MP From ϕ and $\phi \Rightarrow \psi$ infer ψ (Modus Ponens); and

N If ϕ can be obtained without the use of **A6** infer $L\phi$ (Necessitation).

Modus Ponens and a variant of Necessitation are common to all mainstream axiom systems in modal logic. **MP** allows one to perform deductive inference in proving theorems in an axiom system, while **N** implies that an agent can deduce any theorem obtained without the use of **A6** provided he is aware of it. An unusual feature of **N** is the fact that it applies only to formulae that can be derived without **A6**, reflecting the fact that the agent's beliefs concerning $\forall x A x$ are determined by his conjectures. As will be shown below, strengthening **N** to the version that is more common in the literature precludes many of the examples that were used to motivate this approach.

Finally, stronger versions of some of the rationality conditions described above will be used in the characterization results:

PI $L\phi \Longrightarrow LL\phi$;

NI $\neg L\phi \Longrightarrow L\neg L\phi$; and

T $L\phi \Longrightarrow \phi$ for $\phi \in \Phi^{\forall -}$

for any $\phi \in \Phi$. **PI**, **NI**, **T** are difficult to interpret in terms of rationality since the operator L does not represent an epistemic property in its own right. However, it should be noted that, given **A7**, the three conditions respectively imply **PI***, **NI***, and **T***.

To state the main characterization result, write AB for the system comprising the axioms **PC**, **K**, and **A0-7**, and the deductive rules **MP** and **N**. The notation $AB^{\{\mathbf{X}\}}$ then refers to AB supplemented with all the axioms in $\{\mathbf{X}\}$ (so, for example, $AB^{\{\mathbf{T},\mathbf{PI}\}} = AB \cup \mathbf{T} \cup \mathbf{PI}$). Let $\mathcal{D}^r \subset \mathcal{D}$ be the set of structures in which $\mathcal{P}$ is reflexive, $\mathcal{D}^t \subset \mathcal{D}$ be those where $\mathcal{P}$ is transitive, $\mathcal{D}^e$ be those where $\mathcal{P}$ is Euclidean[8], and $\mathcal{D}^b$ be those where $\omega \in \mathcal{P}(\omega')$ implies $\mathcal{A}(\omega) \supseteq \mathcal{A}(\omega')$. The shorthand $\mathcal{D}^{x,y,\dots,z}$ is used for $\mathcal{D}^z \cap \mathcal{D}^{z'} \cap \dots \cap \mathcal{D}^{z''}$.

[8] $\mathcal{P}$ is Euclidean iff $\omega', \omega'' \in \mathcal{P}(\omega)$ implies $\omega' \in \mathcal{P}(\omega'')$.

Theorem 1 *The following hold:*

i. *$\mathcal{D}$ is characterized by AB;*

ii. *$\mathcal{D}^t$ is characterized by $AB^{\mathbf{PI}}$;*

iii. *$\mathcal{D}^e$ is characterized by $AB^{\mathbf{NI}}$;*

iv. *$\mathcal{D}^{r,t,e}$ is characterized by $AB^{\mathbf{PI},\mathbf{NI},\mathbf{T}}$*

v. *$\mathcal{D}^b$ is characterized by $AB^{\mathbf{BA}}$.*

And if $\mathcal{D}^z$, $\mathcal{D}^{z'}$, ..., $\mathcal{D}^{z''}$ are respectively characterized by AB^x, $AB^{x'}$, ..., $AB^{x''}$, then $\mathcal{D}^{z,z',\dots,z''}$ is characterized by $AB^{x,x',\dots,x''}$.

The theorem shows that the rationality conditions can be neatly characterized in terms of properties of $\mathcal{P}$ and $\mathcal{A}$. It should be emphasized that it does not imply (for instance) that there are no structures in $\mathcal{D} \smallsetminus \mathcal{D}^t$ in which **PI** is valid.

Now consider a new subset of $\mathcal{D}$, $\mathcal{D}_T$, containing all those structures where $\mathcal{X}(\omega) = T$ iff $\mathcal{A}(\omega) = P$ for all $\omega, \omega' \in \Omega$. The structures in $\mathcal{D}_T$ are those where the agent's conjecture about whether $\forall x Ax$ is true in any state is always veracious, a feature that renders the first-stage element of the semantics is redundant (that is, it implies $V_1(\omega, \phi) = V(\omega, \phi)$ for all $\omega \in \Omega$ and all $\phi \in \Phi$). The semantics of these structures can be shown to match those described by Halpern and Rego (2009a).

A more conventional version of the derivation rule for necessitation can be expressed as follows:

N′ From ϕ infer $L\phi$.

Use AB^x_T for the axiom system that is identical to AB^x in every respect, except that **N** is replaced by **N′**. The following result, which parallels Theorem 1, can be proved for $\mathcal{D}_T$:

Theorem 2 *The following hold:*

i. *$\mathcal{D}_T$ is characterized by AB_T;*

ii. *$\mathcal{D}^r_T$ is characterized by $AB^{\mathbf{T}}_T$;*

iii. *$\mathcal{D}^t_T$ is characterized by $AB^{\mathbf{PI}}_T$;*

iv. *$\mathcal{D}^e_T$ is characterized by $AB^{\mathbf{NI}}_T$;*

v. *$\mathcal{D}^b_T$ is characterized by $AB^{\mathbf{BA}}_T$.*

And if $\mathcal{D}^x_T$, $\mathcal{D}^y_T$, ..., $\mathcal{D}^z_T$ are respectively characterized by AB^x_T, AB^y_T, ..., AB^z_T, then $\mathcal{D}^{x,y,\dots,z}$ is characterized by $AB^{x,y,\dots,z}_T$.

The principle advantage of using the structures in $\mathcal{D}_T$ is that it allows one to do without the two-stage semantic apparatus – there is no difference between first- and second-stage truth so one may evaluate a formula's truth in any state using the V_1 operator. However, as has been noted by Board and Chung (2007) and Sillari (2008), the structures have the undesirable property that **BA** cannot hold without $\mathbf{P}\forall$ and $\mathbf{BA}, \mathbf{T}, \mathbf{NI}, \mathbf{PI}$ do not hold unless $\mathbf{N}\forall$ does.

Proposition 1 *$\boldsymbol{P}\forall$ is a theorem of $AB_T^{\mathbf{BA}}$ and $\boldsymbol{N}\forall$ is a theorem of $AB_T^{\mathbf{BA},\mathbf{T},\mathbf{NI},\mathbf{PI}}$.*

5 Other Literature and Concluding Remarks

The logical structures defined here draw on the contributions of many other authors. Early pioneering work in this field includes Fagin and Halpern (1988) and Modica and Rustichini (1994 and 1999), who present structures for modeling specific unawareness. Halpern (2001) shows that the latter structures are equivalent to the former when awareness is generated by primitive propositions and **T**, **PI**, and **NI** hold. Thijsse (1996) describes a structure for specific unawareness that assigns truth using two semantic stages, but in other respects his approach is quite different to that proposed here.

Li (2009) and Heifetz et al. (2006) give set-theoretic structures for modeling specific unawareness, the latter with multiple agents. These have been shown, respectively by Heinsalu (2011) and Heifetz et al. (2008), to be equivalent to sub-classes of earlier logical structures of Fagin and Halpern and its multi-agent generalization in Halpern and Rego (2008). A further interactive, set-theoretic approach is provided by Gallanis (2009).

Halpern and Rego (2009a) extend Fagin and Halpern to allow for conscious unawareness, though as noted earlier (1) is inconsistent with (2) in this framework. Board and Chung (2007), Sillari (2008), and Halpern and Rego (2009b) describe structures in which this inconsistency does not arise provided there are impossible states. Board and Chung (2009) render the partitional structures in Board and Chung (2007) in set-theoretic terms; Board et al. (2009) demonstrate that the sub-class of these structures where all states are possible is equivalent to those of Heifetz et al. (2006).

The contribution of this work has been to set out structures where the introspective properties proposed in this literature can be satisfied without introducing impossible states. It is, perhaps, closest in spirit to the dynamic, interactive structures of Grant and Quiggin (2009), where agents' beliefs about the formula "the agent is fully aware" are determined by inductive reasoning, though the current framework is single-agent and static. Walker (2011b) shows how the partitional structures presented here may be translated into set-theoretic terms, using a slightly adapted version of Heifetz et al. (2006). Walker (2011a) then gives a decision theoretic characterization of these set-theoretic structures.

A Proofs

A.1 Proof of Theorem 1

The theorem can be rephrased as claiming that the system AB is *sound* and *complete* with respect to $\mathcal{D}$, where an axiom system is sound wrt some class of structures iff all of its theorems are valid in the class and complete iff every valid formula in the class is a theorem of the system.

A.1.1 Soundness

This amounts to showing that the axioms **PC**, **K**, **A0** – **A7** are valid, and that applications of the deductive rules preserve validity. Some of the arguments below follow a very similar course to well known proofs in, for example, Hughes and Cresswell (1996) and Fagin et al. (1995). I include all of the details because of the non-standard semantics involved here.

PC is valid in virtue of rules O1, O2, and O4 as usual. If **K** were not valid, there would be some a structure such that for state ω (a) $V(\omega, L\phi \wedge L[\phi \Rightarrow \psi]) = 1$ and (b) $V(\omega, L\psi) = 0$. But (a) implies $V_1(\omega', \phi) = 1$ and $V_1(\omega', \phi \Rightarrow \psi) = 1$ for all $\omega' \in \mathcal{P}(\omega)$. It immediately follows that $V_1(\omega', \psi) = 1$ for all $\omega' \in \mathcal{P}(\omega)$, implying $V(\omega, L\psi) = 1$, contradicting (b). **A0** is valid because $\mathrm{Ref}(\forall x A x) = \varnothing$, while **A1** – **A5** are valid in light of the fact that $\mathrm{Ref}(\phi) = \mathrm{Ref}(\neg\phi) = \mathrm{Ref}(L\phi) = \mathrm{Ref}(A\phi) = \mathrm{Ref}(B\phi)$ and $\mathrm{Ref}(\phi) \cup \mathrm{Ref}(\psi) = \mathrm{Ref}(\phi \wedge \psi)$. **A6** follows directly from O2 and O5 and **A7** is implied by O7.

Now suppose ϕ and $\phi \Rightarrow \psi$ are valid formulae. If there were a state ω in any structure such that $V(\omega, \psi) = 0$ then, since ϕ is valid, it must be that $V(\omega, \phi \Rightarrow \psi) = 0$, but this contradicts the fact that $\phi \Rightarrow \psi$ is valid. Therefore ψ is valid and **MP** is validity preserving.

To show that **N** is validity preserving, first introduce the notion of *stage-one validity.* A formula is stage-one valid in D whenever for every state in the state space of D, $V_1(\omega, \phi) = 1$. It is stage-one valid in any class of models $\mathcal{D}^* \subseteq \mathcal{D}$ it is stage-one valid for all $D \in \mathcal{D}^*$. An axiom system is then stage-one sound wrt $\mathcal{D}^*$ iff any theorem ϕ of it is stage-one valid in $\mathcal{D}^*$.

Definition A.1 *The axiom system AB_1^x consists of all of the axioms and deductive rules in AB^x except for the axiom **A6**.*

Lemma A.1 *The axiom system AB_1 is stage-one sound wrt $\mathcal{D}$.*

Proof: The stage-one validity of the axioms **PC**, **K**, **A0-5**, **A7**, and **MP** can be shown in a parallel manner to their validity, proven above.

For **N**, if ϕ is stage-one valid, then it follows immediately that for any ω in any structure, $V(\omega', \phi) = 1$ for all $\omega' \in \mathcal{P}(\omega')$, so $L\phi$ is stage-one valid. Therefore **N** is stage-one validity preserving. $\square$

By definition, the set of theorems of AB_1^x is identical to the set of theorems of AB^x that can be proven without the use of **A7**. Suppose ϕ is a theorem of AB_1, which by Lemma A.1 implies that ϕ is stage-one valid. Hence for every ω in any structure it must be that $V(\omega', \phi) = 1$ for all $\omega' \in \mathcal{P}(\omega)$ and thus that $L\phi$ is valid. **N** is therefore validity preserving.

Thus AB is valid in all of $\mathcal{D}$. To complete the soundness proof for parts (ii)-(v) of the Theorem, proceed as follows:

PI is valid in $\mathcal{D}^t$: If $V(\omega, L\phi) = 1$ for any $\phi \in \Phi$ and in some $D \in \mathcal{D}^t$, then for every $\omega' \in \mathcal{P}(\omega)$, $V_1(\omega', \phi) = 1$. Since $D \in \mathcal{D}^t$, if $\omega'' \in \mathcal{P}(\omega')$ for $\omega' \in \mathcal{P}(\omega)$, then $\omega'' \in \mathcal{P}(\omega')$. Hence if $\omega' \in \mathcal{P}(\omega)$, then for all $\omega'' \in \mathcal{P}(\omega')$, $V_1(\omega'', \phi) = 1$, implying $V_1(\omega', L\phi) = 1$. Since this holds for all $\omega' \in \mathcal{P}(\omega)$, it follows that $V(\omega, LL\phi) = 1$.

NI is valid in $\mathcal{D}^e$: Whenever $V(\omega, \neg L\phi) = 1$ for any $\phi \in \Phi$ and $D \in \mathcal{D}^e$, there is some $\omega' \in \mathcal{P}(\omega)$ such that $V_1(\omega', \neg\phi)$. Since $D \in \mathcal{D}^e$, if $\omega'' \in \mathcal{P}(\omega)$ then $\omega'' \in \mathcal{P}(\omega')$. Therefore if $\omega' \in \mathcal{P}(\omega)$, there is some $\omega'' \in \mathcal{P}(\omega')$ such that $V_1(\omega'', \phi) = 1$, implying $V_1(\omega', \neg L\phi) = 1$. It follows that $V(\omega, L\neg L\phi) = 1$.

T is valid in $\mathcal{D}^r$: First prove that if $\phi \in \Phi^{\forall -}$:

$$V(\omega, \phi) = 1 \implies V_1(\omega, \phi) = 1 \tag{4}$$

Note that for $\phi \in P$ or ϕ of the form $A\psi$, $L\psi$, $B\psi$ where ψ is any member of Φ, the fact that O1 and O5-7 respectively parallel S1 and S5-7 ensure that (4) is satisfied by ϕ. Working inductively, if ψ satisfies (4), then for $\phi = \neg\psi$, $V(\omega, \phi) = 1$ iff $V(\omega, \psi) = 0$ and $V_1(\omega, \phi) = 1$ iff $V_1(\omega, \psi) = 0$, hence ϕ satisfies (4). And if ψ and χ satisfy (4), where $\phi = [\psi \vee \chi]$, $V(\omega, \phi) = 1$ iff at least one of $V(\omega, \psi) = 1$ and $V(\omega, \chi) = 1$ and $V_1(\omega, \phi) = 1$ iff at least one of $V_1(\omega, \psi) = 1$ and $V_1(\omega, \chi) = 1$, so ϕ satisfies (4). Thus, (4) holds for all $\phi \in \Phi^{\forall -}$ as required.

To show that **T** is valid in $\mathcal{D}^r$, suppose $\omega \in \mathcal{P}(\omega)$ and that $V(\omega, L\phi) = 1$ for some $\phi \in \Phi^{\forall -}$ and ω in the state space of some $D \in \mathcal{D}^r$. This requires $V_1(\omega', \phi) = 1$ for all $\omega' \in \mathcal{P}(\omega)$ and therefore (by the reflexivity of $\mathcal{P}$) that $V_1(\omega, \phi) = 1$. By (4) it follows that $V(\omega, \phi) = 1$ and hence that **T** is valid in $\mathcal{D}^r$.

BA is valid in $\mathcal{D}^b$: If $V(\omega, A\phi) = 1$ then $\mathcal{A}(\omega) \supseteq \text{Ref}(\phi)$. For ω in the state space of some $D \in \mathcal{D}^b$, this implies $\mathcal{A}(\omega') \supseteq \text{Ref}(\phi)$ for every $\omega' \in \mathcal{P}(\omega)$ and thus that $V(\omega, LA\phi) = 1$. Since $\text{Ref}(\phi) = \text{Ref}(A\phi)$, this implies $V(\omega, BA\phi) = 1$.

Finally, if **X**, **Y**, ... **Z** are respectively valid in $\mathcal{D}^x$, $\mathcal{D}^y$, ..., $\mathcal{D}^z$, it follows trivially that $\{\mathbf{X}, \mathbf{Y}, \dots, \mathbf{Z}\}$ is valid in $\mathcal{D}^{x,y,\dots,z}$. ■

A.1.2 Completeness

Begin with some preliminary notation and definitions. For any finite set $\Gamma = \{\phi_1, \dots \phi_n\}$, write $\bigwedge\Gamma$ for $[\phi_1 \wedge [\phi_2 \wedge [\dots \phi_n] \dots]$. If Γ is any set of formulae with the property that there is a finite $\Gamma' \subseteq \Gamma$ such that $\bigwedge \Gamma' \Rightarrow \phi$ is a theorem of axiom system AX, write $\Gamma \vdash_{AX} \phi$. The notation $\vdash_{AX} \phi$ then means simply that ϕ is a theorem of of AX.

Say ϕ is *consistent* with axiom system AX – ϕ is AX-consistent – iff $\nvdash_{AX} \neg\phi$ and ϕ is consistent with Γ iff $\Gamma \nvdash_{AX} \neg\phi$. When Γ is a finite set of formulae, it is AX-consistent iff $\nvdash_{AX} \neg \bigwedge \Gamma$, and if Γ is infinite it is AX-consistent iff every finite subset of Γ is AX-consistent. A set of formulae, Γ, is a *maximal AX-consistent set* iff it is consistent and for every $\phi \in \Phi \setminus \Gamma$, $\vdash_{AX} \neg \bigwedge \Gamma \cup \{\phi\}$. Fagin et al. (1995) prove the following lemma:

Lemma A.2 *Provided AX is consistent and contains **PC** and **MP**, if Δ is a AX-consistent set, there exists a maximal AX-consistent set Γ such that $\Gamma \supseteq \Delta$. Furthermore, if Γ is a maximal AX-consistent set, then the following holds:*

i. *For all $\phi \in \Phi$ one of ϕ and $\neg\phi$ is a member of Γ;*

ii. *$[\phi \vee \psi] \in \Gamma$ iff at least one of $\phi \in \Gamma$ and $\psi \in \Gamma$;*

iii. *If ϕ and $\phi \Rightarrow \psi$ are in Γ, then ψ is a member of Γ; and*

iv. *$AX \vdash \phi$ implies $\phi \in \Gamma$.*

Now define an AX-*acceptable* set of formulae as any Γ such that whenever $\Gamma \vdash_{AX} Ap$ for all but finite $p \in P$, $\Gamma \vdash_{AX} \forall x Ax$. A similar property (also called acceptability) was introduced by Halpern and Rego (2009a), and plays the same role in their completeness proof as it does here. Lemmas A.3-A.5 parallel analogous results in Halpern and Rego.

Lemma A.3 *If Γ is finite, then Γ is AB^x-acceptable for any x.*

Proof: Since P is infinite, there is no finite set of formulae, Γ, that does not contain $\forall x Ax$ and that is such that $\Gamma \vdash_{AB^x} Ap$ for all but finite $p \in P$. Therefore, for finite Γ, if $\Gamma \vdash_{AB^x} Ap$ for all but finite $p \in P$, then $\forall x Ax \in \Gamma$ and thus $\Gamma \vdash_{AB^x} \forall x Ax$, so Γ is acceptable. □

Lemma A.4 *If Γ is AB^x-acceptable, then $\Gamma \cup \{\phi\}$ is AB^x-acceptable for any $\phi \in \Phi$.*

Proof: Suppose $\Gamma \cup \{\phi\}$ is not AB^x-acceptable, so $\Gamma \cup \{\phi\} \vdash_{AB^x} Ap$ for all but finite $p \in P$ but not $\Gamma \cup \{\phi\} \vdash_{AB^x} \forall x Ax$. Since Γ is AB^x-acceptable, there must be an infinite number of $p \in P$ such that $\Gamma \nvdash_{AB^x} Ap$ and hence it can only be that $\Gamma \cup \{\phi\} \vdash_{AB^x} Ap$ for all but finite $p \in P$ if $\phi = \forall x Ax$, but this obviously implies $\Gamma \cup \{\phi\} \vdash_{AB^x} \forall x Ax$, a contradiction. □

Lemma A.5 *If Γ is AB^x-consistent and -acceptable, then there exists some maximal AB^x-consistent and -acceptable Γ' containing Γ.*

Proof: Enumerate Φ, $\phi_1, \phi_2 \ldots$ with $\phi_1 = \forall x Ax$ and $\phi_2 = \neg\forall x Ax$. Define $\Delta_0 := \Gamma$. If $\Delta_0 \vdash \neg\forall x Ax$, choose some p such that $\Delta_0 \nvdash_{AB^x} Ap$ (since Δ_0 is AB^x-consistent, $\Delta_0 \nvdash_{AB^x} \forall x Ax$ and thus, as Δ_0 is AB^x-acceptable, there must exist some such p), then let $\Delta_1 := \Delta_0 \cup \{\neg Ap\}$. Otherwise, let $\Delta_1 := \Delta_0 \cup \{\forall x Ax\}$. Note that Δ_1 is always AB^x-consistent and -acceptable.

For $k > 1$, let $\Delta_k = \Delta_{k-1} \cup \{\phi_k\}$ unless $\Delta_{k-1} \vdash \neg\phi_k$ in which case let $\Delta_k = \Delta_{k-1}$. By Lemma A.4, Δ_k is AB^x-acceptable and by construction it is AB^x-consistent.

Define $\Gamma' = \bigcup \Delta_k$ and note that it is maximal AB^x-consistent. To prove that it is also AB^x-acceptable, suppose that for every $p \in P$, $\Gamma \vdash_{AB^x} Ap$. Since Γ' is AB^x-consistent, this means that Δ_1 contains no formula $\neg Ap$, which means that Δ_1 does include $\forall x Ax$. Thus $\Gamma' \vdash_{AB^x} \forall x Ax$ and hence Γ' is AB^x-acceptable. □

For each maximal AB-consistent set, Γ, define two further sets of formulae Γ^T and Γ^F as follows (where $Z \in \{T, F\}$):

$$
\begin{aligned}
p \in \Gamma^Z &\iff p \in \Gamma && (5)\\
\forall x Ax \in \Gamma^T & \qquad\quad \forall x Ax \notin \Gamma^F && (6)\\
\neg\phi \in \Gamma^Z &\iff \phi \notin \Gamma^Z && (7)\\
[\phi \vee \psi] \in \Gamma^Z &\iff \text{At least one of } \phi \in \Gamma^Z \text{ and } \psi \in \Gamma^Z && (8)\\
A\phi \in \Gamma^Z &\iff A\phi \in \Gamma && (9)\\
L\phi \in \Gamma^Z &\iff L\phi \in \Gamma && (10)\\
B\phi \in \Gamma^Z &\iff B\phi \in \Gamma && (11)
\end{aligned}
$$

Lemma A.6 *Where Γ is maximal AB^x-consistent, Γ^T and Γ^F are maximal AB_1^x-consistent.*

Proof: Recall the definition of the set $\Phi^{\forall-}$:

$$
\begin{aligned}
p \in P &\implies p \in \Phi^{\forall-}\\
\text{For any } \phi \in \Phi, & \qquad\quad A\phi, L\phi, B\phi \in \Phi^{\forall-}\\
\phi \in \Phi^{\forall-} &\implies \neg\phi \in \Phi^{\forall-}\\
\phi, \psi \in \Phi^{\forall-} &\implies [\phi \vee \psi] \in \Phi^{\forall-}
\end{aligned}
$$

Note first that for any $\phi \in \Phi^{\forall-}$, if $\phi \in \Gamma$ then $\phi \in \Gamma^T \cap \Gamma^F$. Since $\forall x Ax$ and $\neg\forall x Ax$ are each AB_1^x-consistent with any AB_1^x-consistent subset of $\Phi^{\forall-}$, it follows that $\left(\Gamma \cap \Phi^{\forall-}\right) \cup \{\forall x Ax\}$ and $\left(\Gamma \cap \Phi^{\forall-}\right) \cup \{\neg\forall x Ax\}$ are AB_1^x-consistent sets.

To prove that Γ^T and Γ^F are AB_S^x-maximal consistent sets, I show that for every $\phi \in \Phi$, either ϕ or $\neg\phi$ belongs to each of them. Since Γ is maximal AB_1^x-consistent, by Lemma A.2 it must be that for any $\phi \in \Phi^{\forall-}$, either ϕ or $\neg\phi$ belongs to Γ and hence to Γ^T and Γ^F. Therefore, given rules (6) and (7) above, for any $\phi \in \Phi^{\forall-} \cup \{\forall x Ax, \neg\forall x Ax\}$, it is true of both Γ^T and Γ^F that they contain one of ϕ or $\neg\phi$.

Now note that Φ is simply the closure of $\Phi^{\forall-} \cup \{\forall x Ax, \neg\forall x Ax\}$ under negation and disjunction. Working inductively, if one of ϕ or $\neg\phi$ belongs to Γ^T, then by rule (7) one of $\neg\phi$ and $\neg\neg\phi$ belongs to Γ^T; and if one of ϕ or $\neg\phi$ plus one of ψ and $\neg\psi$ belongs to Γ^T, by rules (8) and (7) one of $[\phi \vee \psi]$ and $\neg[\phi \vee \psi]$. The same argument holds for Γ^F, hence for every $\phi \in \Phi$ one of ϕ and $\neg\phi$ belongs to Γ^T and one belongs to Γ^F. Therefore Γ^T and Γ^F are maximal AB_1^x-consistent. □

Lemma A.7 *If Γ is a maximal AB^x-consistent set, the only maximal AB_1^x-consistent sets containing $\Gamma \cap \Phi^{\forall-}$ are Γ^T and Γ^F.*

Proof: Consider any maximal AB^x-consistent Γ. Clearly $\Gamma \cap \Phi^{\forall-}$ is a maximal AB_1^x-consistent subset of $\Phi^{\forall-}$ and (by Lemma A.2) if $\Gamma_S \supset \Gamma \cap \Phi^{\forall-}$ and Γ_S is maximal AB_1^x-consistent then one of $\forall x Ax$ and $\neg\forall x Ax$ belongs to Γ_S. Therefore any maximal AB_1^x-consistent set containing $\Gamma \cap \Phi^{\forall-}$ also contains a maximal consistent subset of $\left(\Gamma \cap \Phi^{\forall-}\right) \cup \{\forall x Ax, \neg\forall x Ax\}$. There are only two such subsets: $\left(\Gamma \cap \Phi^{\forall-}\right) \cup \{\forall x Ax\}$ and $\left(\Gamma \cap \Phi^{\forall-}\right) \cup \{\neg\forall x Ax\}$. The first is a subset of Γ^T and the second a subset of Γ^F, both of which (by Lemma A.6) are maximal AB_1^x-consistent sets.

Now since Φ is simply the closure of $\Phi^{\forall-} \cup \{\forall x Ax, \neg\forall x Ax\}$ under negation and disjunction, if Δ is a maximal AB_1^x-consistent subset of $\Phi^{\forall-} \cup \{\forall x Ax, \neg\forall x Ax\}$ and ϕ is any formula, either $\Delta \vdash_{AB_1^x} \phi$ or $\Delta \vdash_{AB_1^x} \neg\phi$ (this was demonstrated in the proof of Lemma A.6). Hence if Γ_S is a maximal AB_1^x-consistent set and $\Gamma_S \supseteq \Delta$, $\Delta \vdash_{AB_1^x} \phi$ for all $\phi \in \Gamma_S$. This means that there is only one Γ_S where $\Gamma_S \supseteq \Delta$ and Γ_S is maximal AB_1^x-consistent, which establishes that the only maximal AB_1^x-consistent sets containing $\Gamma \cap \Phi^{\forall-}$ are Γ^T and Γ^F. □

Notation A.1 *From here on in, the following conventions are adopted:*

i. Γ, Γ' *refer generically to maximal AB^x-consistent sets;*

ii. For any Γ, Γ^T and Γ^F refer to the sets defined as in (5)-(11). (Given Lemma A.6 it is taken as read that these sets are maximal AB_1^x-consistent).

iii. Γ^Z, Γ'^Z *refer generically to any maximal AB_1^x-consistent sets defined as in (5)-(11), with Z understood to take either the value T or F.*

Now enumerate the set of all acceptable maximal AB^x-consistent sets $\Gamma_1, \Gamma_2, \ldots$, and use Γ_i^Z to refer to the maximal AB_1^x-consistent sets where $\Gamma_i^Z \cap \Phi^{\forall-} = \Gamma_i \cap \Phi^{\forall-}$. For any set of formulae, Δ, let $\mathrm{Ken}(\Delta)$ refer to $\{\phi : L\phi \in \Delta\}$, and then say Γ_i^Z is *relevant* iff there exists some Γ_j such that $\mathrm{Ken}(\Gamma_j) \subseteq \Gamma_i^Z$.

Lemma A.8 *For every Γ_i, at least one of Γ_i^T and Γ_i^F is relevant in D_c^x.*

Proof: Take any Γ_i, and first show that at least one of Γ_i^T and Γ_i^F is AB^x-consistent. By Lemma A.7, the only two maximal AB_1^x-consistent formulae containing $\Gamma_i \cap \Phi^{\forall-}$ are Γ_i^T and Γ_i^F. Since $\Gamma_i \cap \Phi^{\forall-}$ is AB^x-consistent and can thus (by Lemma A.2) be extended to a maximal AB^x-consistent set, and since AB^x-consistency implies AB_1^x-consistency, it follows that at least one of Γ_i^T and Γ_i^F is AB^x-consistent.

Now observe that there exists a maximal AB^x-consistent set, Γ_j, in which $L\phi$ implies ϕ is a theorem of AB^x. Therefore, if Γ_k is maximal AB^x-consistent, $\text{Ken}(\Gamma_j) \subseteq \Gamma_k$. Therefore, whichever of Γ_i^T and Γ_i^F is maximal AB^x-consistent is also relevant. □

To show completeness, introduce a *canonical model* for AB^x, D_c^x. The state space of the canonical model for AB^x contains one distinct state corresponding to each relevant Γ_i^Z and no other states. Write ω_i^Z for the state that corresponds to Γ_i^Z. The remaining elements of D_c^x are as follows:

$$\begin{aligned} p \in \mathcal{V}(\omega_i^Z) &\iff p \in \Gamma_i \\ \omega_j^{Z'} \in \mathcal{P}(\omega_i^Z) &\iff \text{Ken}(\Gamma_i) \subseteq \Gamma_j^{Z'} \\ p \in \mathcal{A}(\omega_i^Z) &\iff Ap \in \Gamma_i \\ \mathcal{X}(\omega_i^Z) = Z & \qquad\ \ \text{for all } \omega_i^Z \end{aligned}$$

To obtain the result, work in two stages.

Lemma A.9 *For any $\phi \in \Phi$ and any ω in the state space of D_c^x:*

$$V_1(\omega_i^Z, \phi) = 1 \iff \phi \in \Gamma_i^Z \tag{12}$$

Proof: For $\phi \in P$, (12) is trivial. Where $\phi = \forall x A x$, $V_1(\omega_i^Z, \phi) = 1$ iff $\mathcal{X}(\omega_i^Z)) = T$ iff $Y = T$ iff (by (6)) $\phi \in \Gamma_i^Z$.

Proceeding inductively where ψ and χ satisfy (12), then for $\phi = \neg\psi$, S3 says $V_1(\omega_i^Z, \phi) = 1$ iff $V_1(\omega_i^Z, \psi) = 0$, and part (i) of Lemma A.2 implies that $\neg\psi \in \Gamma_i^Z$ iff $\psi \notin \Gamma_i^Z$ iff $V_1(\omega_i^Z, \psi) = 0$. Thus (12) is satisfied for ϕ.

Where $\phi = [\psi \vee \chi]$, S4 says that $V_1(\omega_i^Z, \phi) = 1$ iff at least one of $V_1(\omega_i^Z, \psi) = 1$ and $V_1(\omega_i^Z, \chi) = 1$ iff at least one of $\psi \in \Gamma_i^Z$ and $\chi \in \Gamma_i^Z$. Part (ii) of Lemma A.2 shows that this holds if and only if $\phi \in \Gamma_i^Z$, so (12) is satisfied for ϕ.

If $\phi = A\psi$, $V_1(\omega_i^Z, \phi) = 1$ iff $\mathcal{A}(\omega_i^Z) \supseteq \text{Ref}(\psi)$ iff $Ap \in \Gamma_i^Z$ for all $p \in \text{Ref}(\psi)$ iff (by **A0-5**) $\phi \in \Gamma_i^Z$.

Where $\phi = L\psi$, $\phi \in \Gamma_i^Z$ iff $\psi \in \text{Ken}(\Gamma_i^Z)$ iff $\psi \in \text{Ken}(\Gamma_i)$. This implies that $\psi \in \Gamma_j^{Z'}$ for all $\omega_j^{Z'} \in \mathcal{P}(\omega_i^Z)$. As ψ satisfies (12), it follows that $V_1(\omega_j^{Z'}, \psi) = 1$ for all $\omega_j^{Z'} \in \mathcal{P}(\omega_i^Z)$. By S6, this means $V_1(\omega_i^Z, \phi) = 1$.

Working the other way, if $\phi \notin \Gamma_i^Z$, then $\text{Ken}(\Gamma_i) \cup \{\neg\psi\}$ must be AB_1^x-consistent. To see this, suppose $\text{Ken}(\Gamma_i) \cup \{\neg\psi\}$ were AB_1^x-inconsistent; then there would be some finite set $\{\psi_1, \psi_2, \ldots, \psi_n\} \subseteq \text{Ken}(\Gamma_i)$ such that:

$$\vdash_{AB_1^x} \textstyle\bigwedge_{i=1}^n \psi_i \Rightarrow \psi$$

in which case by **N**, $\bigwedge_{i=1}^n \psi_i \Rightarrow \psi \in \text{Ken}(\Gamma_i)$ and by **K**, $\psi \in \text{Ken}(\Gamma_i)$, a contradiction. Therefore $\text{Ken}(\Gamma_i) \cup \{\neg\psi\}$ is AB_1^x-consistent, so by Lemma A.2 there exists some maximal AB_1^x-consistent $\Gamma_j^{Z'}$ such that $\text{Ken}(\Gamma_i) \cup \{\neg\psi\} \subseteq \Gamma_j^{Z'}$. Since $\text{Ken}(\Gamma_j) \subseteq \Gamma_j^{Z'}$, $\Gamma_j^{Z'}$ is relevant, and hence $\omega_j^{Z'}$ is defined in the state space of D_c^x and $\omega_j^{Z'} \in \mathcal{P}(\omega_i^Z)$. Since ϕ (and hence $\neg\psi$) satisfies (12), it must be that $V_1(\omega_j^{Z'}, \neg\psi) = 1$, in which case $V_1(\omega_i^Z, \phi) = 0$.

Finally, for $\phi = B\psi$, $V_1(\omega_i^Z, B\psi) = 1$ iff $V_1(\omega_i^Z, A\psi) = 1$ and $V_1(\omega_i^Z, L\psi) = 1$ iff $A\psi, L\psi \in \Gamma_i^Z$ iff (by **A7**) $B\psi \in \Gamma_i^Z$. □

Lemma A.10 *For every $\phi \in \Phi$:*

$$V(\omega_i^Z, \phi) = 1 \iff \phi \in \Gamma_i \tag{13}$$

It should be stressed that (13) does *not* require $V(\omega_i^Z, \phi) = 1$ iff $\phi \in \Gamma_i^Z$.

Proof: Where ϕ is a primitive proposition, (13) is satisfied straightforwardly by virtue of the definition of $\mathcal{V}$ above. If $\phi = Ap$ for any $p \in P$, then $\phi \in \Gamma_i$ iff $p \in \mathcal{A}(\omega_i^Z)$, which holds iff $V(\omega_i^Z, \phi) = 1$. For $\phi = \forall x Ax$, if $\phi \in \Gamma_i$ then (by **A6**) $Ap \in \Gamma_i$ for all $p \in P$, implying $V(\omega_i^Z, \phi) = 1$; and if $V(\omega_i^Z, \phi) = 1$ then $V(\omega_i^Z, Ap) = 1$ for all $p \in P$, implying $Ap \in \Gamma_i$ for all $p \in P$, which, given that Γ_i is AB^x-acceptable, implies $\phi \in \Gamma_i$.

Most of the remainder of the proof proceeds inductively in a parallel fashion to that for Lemma A.9. The single point of departure is the argument that if ψ satisfies (13) then $\phi = L\psi$ also satisfies (13). This is now as follows. We have $\phi \in \Gamma_i$ iff $\phi \in \Gamma_i^T$ and $\phi \in \Gamma_i^F$ iff (given Lemma A.9) $V_1(\omega_i^Z, \phi) = 1$ for $Z \in \{T, F\}$ iff $V_1(\omega_j^{Z'}, \psi) = 1$ for all $\omega_j^{Z'} \in \mathcal{P}(\omega_i^Z)$ and $Z \in \{T, F\}$ iff $V(\omega_i^Z, \phi) = 1$ for $Z \in \{T, F\}$. □

Lemma A.8 implies that every acceptable and maximal AB^x-consistent set Γ_i has some state ω_i^Z in the canonical model. Lemmas A.3 and A.5 imply that every AB^x-consistent formula belongs to some acceptable maximal AB^x-consistent set, so therefore (13) implies that every AB-consistent formula is true in some state of the canonical model for AB^x. Therefore the only formulae that are not true in any state of D_c^x are those that are inconsistent – that is, the negations of theorems AB^x – so the only formulae that are true in every state of D_c^x are theorems of AB^x. Therefore any formula that is valid in $\mathcal{D}$ is a theorem of D_c, so AB is complete with respect to $\mathcal{D}$.

To demonstrate that AB^x is complete with respect to $\mathcal{D}^z$, it suffices to show that D_c^x is a member of $\mathcal{D}^z$. For this, proceed as follows:

$D_c^{\mathbf{PI}}$ **belongs to $\mathcal{D}^t$:** If $\omega_j^{Z'} \in \omega_i^Z$, then $\mathrm{Ken}(\Gamma_i) \subseteq \Gamma_j^{Z'}$ and, by **PI**, $\mathrm{Ken}(\Gamma_i) \subseteq \mathrm{Ken}(\Gamma_j^{Z'}) = \mathrm{Ken}(\Gamma_j)$. Hence, if $\omega_k^{Z''} \in \mathcal{P}(\omega_j^{Z'})$ and therefore $\mathrm{Ken}(\Gamma_j) \subseteq \Gamma_k^{Z''}$, $\mathrm{Ken}(\Gamma_i) \subseteq \Gamma_k^{Z'}$ implying $\omega_k^{Z''} \in \mathcal{P}(\omega_i^Z)$. Thus $\mathcal{P}$ is transitive in $D_c^{\mathbf{PI}}$ and therefore $D_c^{\mathbf{PI}} \in \mathcal{D}^t$.

$D_c^{\mathbf{NI}}$ **belongs to $\mathcal{D}^e$:** If $\omega_j^{Z'} \in \mathcal{P}(\omega_i^Z)$, $\mathrm{Ken}(\Gamma_i) \subseteq \Gamma_j^{Z'}$ and by **NI** and part (i) of Lemma A.2 if $\phi \notin \mathrm{Ken}(\Gamma_i)$ then $\phi \notin \mathrm{Ken}(\Gamma_j^Z) = \mathrm{Ken}(\Gamma_j)$. Thus $\mathrm{Ken}(\Gamma_j) \subseteq \mathrm{Ken}(\Gamma_i)$ and hence if $\omega_k^{Z''} \in \mathcal{P}(\omega_i^Z)$ then $\mathrm{Ken}(\Gamma_j) \subseteq \Gamma_k^{Z''}$, implying $\omega_k^{Z''} \in \mathcal{P}(\omega_j^{Z'})$. Thus $\mathcal{P}$ in $D_c^{\mathbf{NI}}$ is Euclidean and therefore $D_c^{\mathbf{NI}} \in \mathcal{D}^e$.

$D_c^{\mathbf{T,PI,NI}}$ **belongs to $\mathcal{D}^r$:** First show that under **T**, $\mathrm{Ken}(\Gamma_i)$ must be $AB_1^{\mathbf{T}}$-consistent for any maximal $AB^{\mathbf{T}}$-consistent Γ_i. For if there was some $\phi \in \mathrm{Ken}(\Gamma_i)$ such that $\vdash_{AB_1^{\mathbf{T}}} \neg\phi$, then it holds by **PC** that for some $\psi \in \Phi^{\forall-}$, $\vdash_{AB_1^{\mathbf{T}}} \phi \Rightarrow [\psi \wedge \neg\psi]$, hence by **N**, $\phi \Rightarrow [\psi \wedge \neg\psi] \in \mathrm{Ken}(\Gamma_i)$ and therefore by **K**, $[\psi \wedge \neg\psi] \in \mathrm{Ken}(\Gamma_i)$.

But then since $[\psi \wedge \neg\psi] \in \Phi^{\forall-}$, by **T** $[\psi \wedge \neg\psi] \in \Gamma_i$, which is impossible as Γ_i is $AB^{\mathbf{T}}$-consistent.

By **T**, it must be that $\left(\mathrm{Ken}(\Gamma_i) \cap \Phi^{\forall-}\right) \subseteq \left(\Gamma_i \cap \Phi^{\forall-}\right)$. And since $\mathrm{Ken}(\Gamma_i)$ is $AB_1^{\mathbf{T}}$-consistent, by Lemma A.2 it must be a subset of some maximal $AB_1^{\mathbf{T}}$-consistent set that is itself a superset of $\Gamma_i \cap \Phi^{\forall-}$. By Lemma A.7, the only two such sets are Γ_i^T and Γ_i^F. Therefore $\mathrm{Ken}(\Gamma_i) \subseteq \Gamma_i^T$ or $\mathrm{Ken}(\Gamma_i) \subseteq \Gamma_i^F$. So for all ω_i^Z, at least one of ω_i^T and ω_i^F belongs to $\mathcal{P}(\omega_i^Z)$.

To show the result, demonstrate that if $\mathrm{Ken}(\Gamma_i) \nsubseteq \Gamma_i^Z$ then Γ_i^Z is not relevant, and therefore that ω_i^Z is not defined, from which it follows that whenever ω_i^Z is defined, $\omega_i^Z \in \mathcal{P}(\omega_i^Z)$. Suppose this is not the case and that $\mathrm{Ken}(\Gamma_i) \nsubseteq \Gamma_i^Z$ and ω_i^Z is relevant, implying that there is some Γ_j such that $\mathrm{Ken}(\Gamma_j) \subseteq \Gamma_i^Z$. Then by **PI** and **NI**, $\mathrm{Ken}(\Gamma_j) = \mathrm{Ken}(\Gamma_i)$, so $\mathrm{Ken}(\Gamma_i) \subseteq \Gamma_i^Z$, a contradiction.

$D_c^{\mathbf{AB}}$ **belongs to** $\mathcal{D}^b$**:** Suppose $\omega_j^{Z'} \in \mathcal{P}(\omega_i^Z)$ and thus that $\mathrm{Ken}(\Gamma_i) \subseteq \Gamma_j^{Z'}$. If $p \in \mathcal{A}(\omega_i^Z)$ then $Ap \in \Gamma_i^Z$ and by **AB**, $Ap \in \mathrm{Ken}(\Gamma_i)$ and thus $Ap \in \Gamma_j^{Z'}$, implying $p \in \mathcal{A}(\omega_j^{Z'})$. Therefore $\omega_j^{Z'} \in \mathcal{P}(\omega_i^Z)$ implies $\mathcal{A}(\omega_i^Z) \subseteq \mathcal{A}(\omega_j^{Z'})$ so $D_c^{\mathbf{AB}}$ belongs to $\mathcal{D}^b$.

To complete the proof, the arguments above can be used to show that the canonical model of the system $AB^{x,x',\ldots,x''}$ lies within $\mathcal{D}^{z,z',\ldots,z''}$, so $AB^{x,x',\ldots,x''}$ is complete with respect to $\mathcal{D}^{z,z',\ldots,z''}$. ■

A.2 Proof of Theorem 2

If $D \in \mathcal{D}_T$, then it must be that:

$$V(\omega, \phi) = 1 \iff V_1(\omega, \phi) = 1 \tag{14}$$

for any $\phi \in \Phi$ and any ω in D's state space. (14) implies that $V(\omega, L\phi) = 1$ iff $V(\omega', \phi) = 1$ for all $\omega' \in \mathcal{P}(\omega)$. It is then easy to see that $\mathcal{D}_T$ is essentially the same as the structures $\mathcal{M}^{gpp}$ for the language $\mathcal{L}^{KXA}$ described by Halpern (2001), who proves an analogous theorem. ■

A.3 Proof of Proposition 1

Given Theorem 2, it suffices to show the following:

$$\mathbf{P}\forall \text{ is valid in } \mathcal{D}_T^b \tag{15}$$

$$\mathbf{N}\forall \text{ is valid in } \mathcal{D}_T^{r,t,e,b} \tag{16}$$

To show (15), consider any $D \in \mathcal{D}_T^b$ whose state space includes ω. $V(\omega, \forall x Ax) = 1$ implies $\mathcal{A}(\omega') = P$ for all $\omega' \in \mathcal{P}(\omega)$, which given (14) requires $V(\omega, L\forall x Ax) = 1$ and therefore $V(\omega, B\forall x Ax) = 1$.

For (16), take $D \in \mathcal{D}_T^{r,t,e,b}$ with state ω. Since $\mathcal{P}$ is partitional, $\omega' \in \mathcal{P}(\omega)$ implies $\mathcal{A}(\omega) = \mathcal{A}(\omega')$. Therefore $V(\omega, \neg\forall x Ax) = 1$ implies $V(\omega', \neg\forall x Ax) = 1$ for all $\omega' \in \mathcal{P}(\omega)$, which by (14) implies $V(\omega, B\neg\forall x Ax) = 1$. □

References

BOARD, O. J. AND K.-S. CHUNG (2007): "Object-Based Unawareness: Axioms," *mimeo: University of Minnesota.*

——— (2009): "Object-Based Unawareness: Theory and Applications," *mimeo: University of Minnesota.*

BOARD, O. J., K.-S. CHUNG, AND B. C. SCHIPPER (2009): "Two models of unawareness: Comparing the object-based and the subjective-state-space approaches," *mimeo: University of Minnesota.*

DEKEL, E., B. LIPMAN, AND A. RUSTICHINI (1998): "Standard State-Space Models Preclude Unawareness," *Econometrica*, 66, 159–173.

FAGIN, R. AND J. Y. HALPERN (1988): "Belief, Awareness, and Limited Reasoning," *Artificial Intelligence*, 34, 39–76.

FAGIN, R., J. Y. HALPERN, Y. MOSES, AND M. Y. VARDI (1995): *Reasoning about Knowledge*, MIT.

GALLANIS, S. (2009): "Unawareness of Theorems," *mimeo: University of Southampton.*

GRANT, S. AND J. QUIGGIN (2009): "Inductive reasoning about unawarness," *mimeo: Rice University.*

HALPERN, J. Y. (2001): "Alternative Semantics for Unawareness," *Games and Economic Behavior*, 37, 321–339.

HALPERN, J. Y. AND L. C. REGO (2008): "Interactive Unawareness Revisited," *Games and Economic Behavior.*

——— (2009a): "Reasoning about knowledge of unawareness," *Games and Economic Behavior.*

——— (2009b): "Reasoning about knowledge of unawareness revisited," *mimeo: Cornell University.*

HEIFETZ, A., M. MEIER, AND B. C. SCHIPPER (2006): "Interactive Unawareness," *Journal of Economic Theory*, 130, 78–94.

——— (2008): "A canonical model for interactive unawareness," *Games and Economic Behavior*, 62, 304–324.

HEINSALU, S. (2011): "Equivalence of the Information Structure with Unawareness to the Logic of Unawareness," *mimeo: Yale University.*

HUGHES, G. E. AND M. J. CRESSWELL (1996): *A New Introduction to Modal Logic*, Routledge, 2nd ed.

KARNI, E. AND M.-L. VIERO (2009): ""Reverse Bayesianism: a choice-based theory of growing awareness," *mimeo: John Hopkins Univeristy.*

Li, J. (2008): "A Note on Unawareness and Zero Probability," *PIER Working Paper 08-022*.

——— (2009): "Information Structures with Unawareness," *Journal of Economic Theory*, 121, 167–191.

Lipman, B. L. (2003): "Decision Theory without Logical Omniscience: Toward an Axiomatic Framework for Bounded Rationality," *The Review of Economic Studies*.

Modica, S. and A. Rustichini (1994): "Awareness and Partitional Information Structures," *Theory and Decision*, 37, 107–124.

——— (1999): "Unawareness and Partitional Information Structures," *Games and Economic Behavior*, 27, 265–298.

Morris, S. (1996): "The Logic of Belief and Belief Change: A Decision Theoretic Approach," *Journal of Economic Theory*, 69.

Savage, L. J. (1954): *The Foundations of Statistics*, Dover.

Schipper, B. C. (2010): "Revealed Unawareness," *mimeo: Northwestern University*.

Sillari, G. (2008): "Quantified logic of awareness and impossible possible worlds," *Journal of Symbolic Logic*, 1, 1–16.

Thijsse, E. (1996): "Combining partial and classical semantics. A hybrid approach to belief and awareness," in *Partiality, modality, and nonmonotonicity*, ed. by P. Doherty, Stanford: CSLI.

Tirole, J. (2009): "Cognition and Incomplete Contracts," *American Economic Review*, 99, 265–294.

Walker, O. J. (2011a): "Reasoning about Unawareness: a Decision Theoretic Account," *DPhil chapter: University of Oxford*.

——— (2011b): "A Subjective State Space Framework for Conscious Unawareness," *DPhil chapter: University of Oxford*.

Walker, O. J. and S. Dietz (2011): "A Representation Theorem for Choice under Conscious Unawareness," *mimeo: London School of Economics*.

Centre for Climate Change Economics and Policy

Grantham Research Institute on Climate Change and the Environment

Unawareness with 'possible' possible worlds

Oliver Walker

December 2011

Centre for Climate Change Economics and Policy
Working Paper No. 78
Grantham Research Institute on Climate Change and the Environment
Working Paper No. 69

The Grantham Foundation

ESRC Economic & Social Research Council

UNIVERSITY OF LEEDS

LSE The London School of Economics and Political Science

The Centre for Climate Change Economics and Policy (CCCEP) was established by the University of Leeds and the London School of Economics and Political Science in 2008 to advance public and private action on climate change through innovative, rigorous research. The Centre is funded by the UK Economic and Social Research Council and has five inter-linked research programmes:

1. Developing climate science and economics
2. Climate change governance for a new global deal
3. Adaptation to climate change and human development
4. Governments, markets and climate change mitigation
5. The Munich Re Programme - Evaluating the economics of climate risks and opportunities in the insurance sector

More information about the Centre for Climate Change Economics and Policy can be found at: http://www.cccep.ac.uk.

The Grantham Research Institute on Climate Change and the Environment was established by the London School of Economics and Political Science in 2008 to bring together international expertise on economics, finance, geography, the environment, international development and political economy to create a world-leading centre for policy-relevant research and training in climate change and the environment. The Institute is funded by the Grantham Foundation for the Protection of the Environment, and has five research programmes:

1. Use of climate science in decision-making
2. Mitigation of climate change (including the roles of carbon markets and low-carbon technologies)
3. Impacts of, and adaptation to, climate change, and its effects on development
4. Governance of climate change
5. Management of forests and ecosystems

More information about the Grantham Research Institute on Climate Change and the Environment can be found at: http://www.lse.ac.uk/grantham.