A reasoning approach to knowledge, introspection and unawareness*

Olivier Gossner† and Elias Tsakas‡

July 24, 2009

Abstract

We study the knowledge of a reasoning agent who assumes consciousness of all primitive propositions: He thinks that, for each such proposition, he either knows it and knows that he knows it, or doesn't know it and knows that he doesn't know it.

If the agent is really conscious of all primitive propositions, we show that the agent is actually conscious of all propositions, in which case positive and negative introspection hold for all propositions.

If the agent is not conscious of all primitive propositions, but thinks he is, we show that the agent is necessarily unaware of some proposition, or exhibits delusion about his own knowledge.

KEYWORDS: Knowledge, Information, Reasoning, Unawareness.

JEL CLASSIFICATION: D80, D83, D89.

*The financial support from the Adlerbertska Forskningsstiftelsen Foundation is gratefully acknowledged.

†Paris School of Economics, and London School of Economics and Political Science ogossner@ens.fr

‡Department of Economics, Maastricht University e.tsakas@maastrichtuniversity.nl

1 Introduction

The fundamental question of finding a model that describes the knowledge of an economic agent – whether fully rational or boundedly rational – is, many decades after it was raised in the seminal work of Simon (1955), still open to debate.

The requirements that a satisfactory model should comply with are quite demanding, and perhaps too much so. Ideally, such a model would be tractable, and allow to fully describe the whole knowledge of the agent using a limited number of parameters. At the same time, this model should be able to capture not only the features of knowledge of a rational agent, but also of a boundedly rational agent, such as unawareness.

Two major strands of literature have developed relying on semantic models, or on syntactic models.

In a semantic – or else state space – model, the player's knowledge is described by a possibility correspondence which assigns, to each possible state, the set of states that are considered possible by the agent. Semantic models are introduced by Hintikka (1962), Aumann (1976) and Geanakoplos (1989). The case in which the agent's possibility correspondence defines a partition of the state space is a benchmark for the rational agent, but other non-partitional structures have been used as well, and describe interesting properties of a boundedly rational agent's knowledge. A semantic model naturally induces a knowledge operator: An event – subset of the set of states – is known to the agent whenever it is a superset of the set of states the agent considers as possible.

An important real life phenomenon that is desirable to capture in a knowledge model is awareness – or its lack thereof, unawareness. An agent who is unaware of an event E ignores this phenomenon, thus he does not know E. Furthermore, unawareness entails unawareness of one's own ignorance: an agent unaware of E, cannot know that he doesn't know E, and so on. In an important paper, Dekel et al. (1998) prove an impossibility result, showing that any rich enough notion of unawareness cannot be captured by a semantic model.

The basis of a syntactic model (Chellas, 1980; Aumann, 1999) are propositions. The propositions that do not involve the agent's knowledge, such as "it is raining and Ann has blue eyes", are called primitive propositions. Other propositions, such as "the agent lives in New York or does not know that Ann has blue eyes", are called epistemic propositions. A state describes, for each possible proposition, whether this proposition is true or not. In principle, nothing relates the truth value of different propositions at a given state, but in practice, one restricts attention to states fulfilling some logical axioms such as "a proposition and its negation do not hold".

Which set of axioms yields a model that describes the agent's knowledge in a satisfactory manner is a question that does not have a straightforward answer. Weak axioms leads to a very large

number of states, thus to an intractable model, and leaves open the possibility for counter-intuitive situations. On the other hand, stronger axioms preclude the description of interesting features of bounded rationality.

Axioms that play a central role in syntactic models are the axioms of introspection. According to positive introspection, written K_2, whenever the agent knows a proposition, he also knows that he knows this proposition. Negative introspection, written K_3, is the property that the agent knows his own ignorance: Whenever he does not know a proposition, he also knows that he does not know this proposition. An additional natural axiom is non-delusion, written K_1: It says that any proposition known to the agent must be true.

The question that naturally arises is whether introspection is a reasonable assumption. In particular, the mental process through which introspection is achieved by the agent is not explicit. Introspection is assumed to be a cognitive capacity per se, i.e., the agent does not reach introspection through a natural deductive process.

Samet (1990) showed that the combination of non-delusion and introspection has strong consequences: Under $K_1 - K_3$, the agent's knowledge in any agent in a syntactic model can equivalently be described by a partition on a standard state space. One implication of this result is no bounded rationality phenomenon such as unawareness can be captured by a syntactic model with non-delusion and introspection.

It follows from the previous discussion that introspection, although central in syntactic models, appears to be doubly problematic. On the one hand, it may be difficult to justify. On the other hand, it excludes the possibility for interesting ways of bounded rationality.

The approach we take in this paper is to assume that the agent's knowledge has to be obtained either through direct observation, or elaborated through a process of deductive reasoning. Justifying knowledge as much as possible through reasoning rather than direct observation is fundamental in part because we do not wish to assume that higher order knowledge is directly observed by the agent, but aim at explaining it through a deductive process.

Consider the situation of an agent who has prior experience of a phenomenon described by a primitive proposition such as "it is raining". It is reasonable to assume that the agent has a good knowledge of his own knowledge of such propositions, hence neither positive nor negative introspection are overly problematic for such proposition in this case. As a shorthand, we say that the agent is conscious of some proposition when he either knows it and knows that he knows it, or does not know it and knows that he does not know it.

One of our key assumptions is that he also knows, or just assumes, that he is conscious of all primitive propositions. This assumption on the part of the agent may be well founded, in case he is

indeed conscious of all primitives, or ill founded, when there exists a primitive proposition the agent isn't conscious of. We say in the first case that the agent rightly assumes consciousness of primitives, and in the other case that the agent wrongly assumes consciousness of primitives. We study both cases, in which the agent may or may not be conscious of all propositions.

Our main result, Theorem 1, shows that a reasoning agent who rightly assumes consciousness of primitives is necessarily conscious of all propositions. Theorem 1 thus shows that consciousness of propositions extends from the primitive propositions to the full set of propositions. The proof of Theorem 1 is not immediate, and hinges on Theorem 2 which shows that a reasoning agent who assumes consciousness of primitive propositions also *de facto* assumes consciousness of all propositions.

Theorem 1 provides a foundation for introspection in the sense that it allows to decompose positive and negative introspection into smaller elements which each have a clear interpretation 1) the agent is capable of reasoning, 2) the agent assumes consciousness of primitives, and 3) this assumption is well founded.

The knowledge of an agent who wrongly assumes consciousness of primitives is an interesting object of study and provides remarkable insights on bounded rationality phenomena. The results we obtain in this case depend on whether the agent assumes that every proposition that he knows is necessarily true, which we summarize by saying that the agent assumes non-delusion.

Consider a reasoning agent who assumes non-delusion, and who wrongly assumes consciousness of primitives. Then, either positive or negative introspection must fail for some primitive proposition ϕ. We show that, in case of failure of negative introspection, the agent is necessarily unaware of ϕ in a strong sense: He does not know ϕ, does not know that he does not know ϕ, and does not know that ... that he does not know ϕ, where "..." can contain any chain of "knows" and "does not know". Thus, the agent shows a form of complete lack of recognition of ϕ which captures the idea that one has about awareness. In case of a failure of positive introspection, we show that the agent is necessarily unaware of his knowledge of ϕ, where the notion of unawareness is defined in the same strong sense as above.

Therefore for an agent assuming non-delusion, and wrongly assuming consciousness of primitives, unawareness is not only a possibility in the model, but it arises as the only outcome that is compatible with any failure of introspection, whether on a primitive proposition or not.

If we relax the assumption that the agent assumes non-delusion, the two possibilities of unawareness as above occur, as well as two extra possibilities. In the first one, the agent knows some primitive proposition, while thinking that he does not know this primitive. We then say that the agent exhibits delusion on negative knowledge. In the other possibility, the agent does not know some primitive proposition, while thinking that he knows this primitive. We refer to this situation as delusion on

positive knowledge.

Positive or negative delusion are interesting possibilities that have not, to the best of our knowledge, been exploited in the literature. Those situations arise when the agent does not assume non-delusion and wrongly assumes consciousness of primitives.

To summarize, we study the knowledge of a reasoning agent who assumes consciousness of primitives. When this assumption is well founded, the agent is conscious of all propositions. We see this first result as a possible foundation for the introspection axioms. When the agent is not conscious of some primitive, our model gives rise to interesting bounded-rationality features such as unawareness.

2 Knowledge, reasoning, and consciousness

2.1 Propositions

We recall the syntactic model of knowledge from Aumann (1999), Chellas (1980), Fagin et al. (1995). We start with a set of primitive propositions, Φ_0, which describe facts about the world that do not involve the agent's knowledge. Examples of primitive propositions are "it is raining" or "Ann has blue eyes".

The symbols $\neg$, $\vee$ and $\wedge$ express negation, disjunction and conjunction, i.e., $\neg\phi$ stands for "not ϕ", $(\phi_1 \vee \phi_2)$ stands for "ϕ_1 or ϕ_2" and $(\phi_1 \wedge \phi_2)$ stands for "ϕ_1 and ϕ_2". The set of primitive propositions Φ_0 is closed under these operations: $\phi_1 \vee \phi_2$, $\phi_1 \wedge \phi_2$ and $\neg\phi_1$ are primitive propositions whenever ϕ_1, ϕ_2 are.

The symbol K expresses knowledge: $K\phi$ stands for "the agent knows ϕ". Let $B_0(\phi) := \{\phi\}$ and iteratively define $B_n(\phi) := \{K\phi', \neg K\phi' | \phi' \in B_{n-1}(\phi)\}$. Then, define the set of propositions epistemically derived or generated by ϕ as $B(\phi) := \bigcup_{n\geq 0} B_n(\phi)$.

The set of propositions Φ is the closure of Φ_0 under K, $\vee$, $\wedge$ and $\neg$. It is the smallest set of propositions that can be constructed from Φ_0 using these operations. For instance, $K(\phi_1 \vee \phi_2) \wedge K\neg\phi_3$ is a proposition – although non primitive – whenever ϕ_1, ϕ_2 and ϕ_3 are.

The proposition "ϕ_1 implies ϕ_2" is denoted $(\phi_1 \to \phi_2)$ and is an abbreviation for $(\neg\phi_1 \vee \phi_2)$; "ϕ_1 if and only if ϕ_2" is denoted by $(\phi_1 \leftrightarrow \phi_2)$ and is defined as $(\phi_1 \to \phi_2) \wedge (\phi_2 \to \phi_1)$.

2.2 States of the world

A state ω assigns a truth value to every proposition in Φ. It is a mapping from Φ to $\{0, 1\}$, with the interpretation that ϕ is true at ω when $\omega(\phi) = 1$ and false otherwise. We identify ω with the set of propositions that are true at ω, and we write $\phi \in \omega$ when ϕ is true at ω. Thus, we write

$\omega = \{\phi \in \Phi : \omega(\phi) = 1\}$. A state space Ω is a collection of such states ω.

We make some minimal requirements on the truth values of different propositions at states we consider: Let Ω_0 represent the set of states for which the following conditions hold for two arbitray propositions ϕ_1 and ϕ_2:

(A_1) $\phi_1 \in \omega$, if and only if $\neg\phi \notin \omega$,

(A_2) $(\phi_1 \wedge \phi_2) \in \omega$, if and only if $\phi_1 \in \omega$ and $\phi_2 \in \omega$,

(A_3) $(\phi_1 \vee \phi_2) \in \omega$, if and only if $\phi_1 \in \omega$ or $\phi_2 \in \omega$,

(A_4) $K(\phi_1 \wedge \phi_2) \in \omega$, if and only if $K\phi_1 \in \omega$ and $K\phi_2 \in \omega$,

(A_5) if $K\phi_1 \in \omega$ or $K\phi_2 \in \omega$, then $K(\phi_1 \vee \phi_2) \in \omega$,

(A_6) if $K\phi \in \omega$, then $\neg K\neg\phi \in \omega$.

Conditions $A_1 - A_3$ require that states agree with the interpretation we form of the operators of negation ($\neg$), conjunction ($\wedge$) and disjunction ($\vee$), whereas $A_4 - A_6$ are basic requirements on the logical structure of the agent's knowledge: the conjunction of two propositions is known if and only if the two propositions are known, and for the disjunction of two propositions to be known it is sufficient that one of these propositions is known, and both a proposition as its contrary cannot be simultaneously known.

A property that plays a central role in the literature of epistemic knowledge is introspection. Recall that positive introspection holds for ϕ at ω when $K\phi \in \omega$ implies $KK\phi \in \omega$. Negative introspection holds for ϕ at ω whenever $\neg K\phi \in \omega$ implies $K\neg K\phi \in \omega$. Positive (negative) introspection holds at ω whenever positive (negative) introspection holds for all ϕ at ω. Introspection is not viewed as a reasoning process, but is rather assumed to be a separate process, through which the agent is capable of knowing his own knowledge. Although many important results rely on the assumptions of positive and negative introspection, it has often been recognized that these properties lack some form of fundamental justification.

We say that the agent is conscious of a proposition ϕ at state ω when the properties defining positive and negative introspection hold for ϕ at ω: This is the case if either both $K\phi$ and $KK\phi$ are in ω, or both $\neg K\phi$ and $K\neg K\phi$ are in ω.

We let $C\phi$ be an abbreviation for $(K\phi \wedge KK\phi) \vee (\neg K\phi \wedge K\neg K\phi)$, and we read $C\phi$ as "the agent is conscious of ϕ". The proposition $C\phi$ can be rewritten as $(K\phi \rightarrow KK\phi) \wedge (\neg K\phi \rightarrow K\neg K\phi)$.

We do not wish to assume that the agent is conscious of every proposition (hence imposing the demanding requirements of positive and negative introspection), but rather aim to analyze how the

agent can, through a reasoning process, extend consciousness from a restricted set of propositions to the whole set of propositions. We say that the agent is conscious of primitives at state ω when the agent is conscious of every primitive proposition at ω, and that the agent is fully conscious at ω when the agent is conscious of every proposition at ω.

Obviously, consciousness of primitives is a much weaker requirement than full consciousness. Full consciousness seems difficult to justify, since it requires the agent to be capable of both positive and negative introspection. On the other hand, if the agent has some familiarity with the environment he lives in, it is entirely conceivable that he is conscious of primitive propositions. Consciousness of primitive propositions can even be obtained in a "hard wired" manner if, upon receiving a stimulus corresponding to the observation of some primitive ϕ, the agent also receives a signal corresponding to the information "ϕ is known".

2.3 Reasoning

The agent's knowledge is built both on direct observation and trough reasoning. This reasoning allows to derive knowledge from directly observed knowledge, but also to infer propositions that are known independently of any directly observed knowledge from other such propositions. The propositions which are always assumed to be true by the agent are called tautologies, and are the object of this section.

The set of tautologies is a subset T of the set Φ of propositions. We make the following assumptions on the set of tautologies. For every $\phi_1, \phi_2 \in \Phi$

(A'_0) All tautologies of propositional calculus are in T

(A'_1) $\big(K(\phi_1 \wedge \phi_2) \leftrightarrow (K\phi_1 \wedge K\phi_2)\big) \in T$,

(A'_2) $\big((K\phi_1 \vee K\phi_2) \rightarrow K(\phi_1 \vee \phi_2)\big) \in T$,

(A'_3) $(K\phi_1 \rightarrow \neg K\neg\phi_1) \in T$.

Note that the properties listed in $A'_0 - A'_3$ are assumed to be true in every state of the world, so that is it natural to assume that the agent knows these rules.

As part of his reasoning process, the agent is capable of deriving tautologies from other tautologies. We therefore assume:

(R_1) $(\phi_1 \wedge \phi_2) \in T$, if and only if $\phi_1 \in T$ and $\phi_2 \in T$,

(R_2) if $\phi_1 \in T$ and $(\phi_1 \rightarrow \phi_2) \in T$, then $\phi_2 \in T$ (Modus Ponens).

Modus Ponens requires that the agent is capable of making inferences on the set of tautologies. The rule R_1 is close in spirit, it requires that the conjunction of two tautologies, is also a tautology.

Furthermore, we require the agent's knowledge to satisfy the following standard rule of reasoning, which states that knowing a proposition and also knowing the implications of this proposition, yields knowledge of the implications:

(R_I) if $(\phi_1 \to \phi_2) \in T$, then $(K\phi_1 \to K\phi_2) \in T$ (rule of inference).

Finally, as we have in mind an agent who knows all tautologies, the knowledge of these tautologies is part of the tautologies themselves:

(R_T) if $\phi \in T$ then $K\phi \in T$ (rule of necessitation),

All the assumptions $A'_0 - A'_3$, $R_1 - R_2$, R_I and R_T are common when describing a reasoning agent, as in the case of a logically omniscient agent (see e.g., Chellas, 1980).

In the remainder of the paper, when referring to a reasoning agent, we have in mind the assumption that the set T of tautologies satisfies the axioms $A'_0 - A'_3$, $R_1 - R_2$, R_I and R_T.

The following assumption is also common while describing a logically omniscient agent, such as in the system S5 of modal logic.

(R_C) $C\phi \in T$, for every $\phi \in \Phi$.

The rule R_C can be understood as "the agent believes that he is conscious of all propositions".

There is *a priori* no reason for the agent to assume that he is conscious of all propositions. On the other hand, if every state has the property that, at this state, that the agent is conscious of all primitives, it is more natural to make the weaker following assumption on the set of tautologies:

(R_{CP}) $C\phi \in T$, for every $\phi \in \Phi_0$.

To refer to R_{CP}, or R_C respectively, we say that the agent assumes consciousness of primitives, or assumes full consciousness.

We let T_C respectively T_{CP} denote be the minimal set of tautologies which satisfies $A'_0 - A'_3$, $R_1 - R_2$, R_I, R_T and R_C (respectively R_{CP}).

3 A conscious agent state space

We come back to the state space model, and consider a reasoning agent whose tautologies are described by T_C.

To the axioms describing the logical consistency describing a state in Ω_0, we add the assumptions that the agent knows the tautologies, and is capable of making logical deductions. Then, we let Ω_U be the set of states $\omega \in \Omega_0$ such that:

(K_T) $K\phi \in \omega$, for every $\phi \in T$, and

(K_I) if $K(\phi_1 \to \phi_2) \in \omega$, then $(K\phi_1 \to K\phi_2) \in \omega$, for every $\phi_1, \phi_2 \in \Phi$.

In the state space Ω_U, no assumption is made on the consciousness of the agent about any propositions. We let Ω_C be the subset of states in Ω_U where the agent is conscious of all primitive propositions:

(K_C) $C\phi \in \omega$, for every $\phi \in \Phi_0$.

To refer to a state in Ω_C, we say that the agent rightly assumes that he is conscious of primitive propositions since he both makes this assumption, and this assumption is true.

We now state our second result.

Theorem 1. *If the reasoning agent rightly assumes consciousness of primitive propositions, he is conscious of all propositions: For all* $\omega \in \Omega_C$

$$C\phi \in \omega, \text{ for every } \phi \in \Phi.$$

Theorem 1 provides a foundation for the rules of introspection, which play a central role in the description of agent's knowledge. It requires three assumptions. The first is that the agent has sufficient understanding of his environment, in the sense that he is conscious of all the primitive propositions that arise in this environment. The second is that the agent assumes that he is conscious of these primitive propositions, in the sense that consciousness of primitives is part of the agent's tautologies. Finally, the third assumption is that the agent is capable of reasoning, both on the set of tautologies and in making inferences from directly observed knowledge. Under these three assumptions, both positive and negative introspection hold for every propositions.

We find this conclusion to have a strong impact on the complexity of the description of a state, as well as on the cardinality of the state space, to be discussed later.

The proof of Theorem 1 is a quite direct consequence of the next result, which concerns the set of tautologies of an agent who assumes consciousness of primitives.

Theorem 2. *If the reasoning agent assumes consciousness of primitive propositions, he also assumes full consciousness, i.e.,*

$$T_C = T_{CP}.$$

The proof of Theorem 2 is rather long and involved. It can be found in appendix.

Proof of Theorem 1 from Theorem 2. The tautologies in T_C which are in $A'_0 - A'_3$ are satisfied in any $\omega \in \Omega_0$. Tautologies of the form T_C are true in every state $\omega \in \Omega_C$. The rules according to which tautologies are derived correspond to requirements on the states in Ω_C. □

4 Unawareness and knowledge delusion

The state space Ω_C describes the knowledge of an agent who knows everything about his own knowledge. The foundation of this knowledge is the consciousness of primitives, while the agent also assumes the same consciousness of primitives.

We find that although the state space Ω_C can be adequate to describe the agent's knowledge in many instances, it can fail to be rich enough to encompass situations which exhibit bounded rationality failures, such as for instance unawareness.

In its generally accepted meaning, unawareness takes place when the agent fails to recognize some proposition, and even also fails to recognize this failure itself. This is the situation of, for instance, the buyer of a house who doesn't know that, prior to buying the house, the possibility of termites in the house is an event to be checked, against which no law protection would exist should termites be found after the sale takes place.

We take the position that awareness, or unawareness, cannot distinguish the agent using his set of tautologies. On the other hand, what may distinguish them is their consciousness of primitive propositions.

The aim of this section is to study the more general state space Ω_U where the reasoning agent assumes to be conscious of all primitive propositions – and therefore of all propositions – but he may or may not be conscious of these.

The states in Ω_C where the agent is indeed conscious of all primitive propositions has been studied in Section 3. When considering a state ω that belongs to Ω_U but not to Ω_C, we say that the agent wrongly assumes that he is conscious of primitives (at ω).

In order to study the structure of knowledge of an agent who wrongly assumes consciousness of primitives, we need to express some result about the set of tautologies T_C, which are assumed to be true by the agent.

Consider any sequence $\mathbf{K} = \tau_1, \ldots, \tau_n$, $n \geq 1$, where for every $i = 1, \ldots, n$, $\tau_i = K$ or $\tau_i = \neg K$. For such a sequence $\mathbf{K}$, we define its parity $p(\mathbf{K}) \in \{0, 1\}$ as the parity of the number of occurrences of $\neg K$ in $\mathbf{K}$. For instance, $p(\neg K K \neg K) = p(K) = 0$, whereas $p(K \neg K) = p(\neg K \neg K \neg K) = 1$, i.e., $p(\mathbf{K}) = 0$ if the number of negations in $\mathbf{K}$ is even, and $p(\mathbf{K}) = 1$ otherwise.

Lemma 1. *Consider a reasoning agent. For any two sequences* $\mathbf{K}$ *and* $\mathbf{K}'$ *such that*

$$p(\mathbf{K}) = p(\mathbf{K}')$$

and for any proposition ϕ *we have*

$$(\mathbf{K}\phi \leftrightarrow \mathbf{K}'\phi) \in T.$$

In particular, for every $\omega \in \Omega_U$*,*

$$K(\mathbf{K}\phi) \in \omega, \text{ if and only if } K(\mathbf{K}'\phi) \in \omega.$$

Proof. It follows from Theorem 2 that $C\phi \in T$ for all $\phi \in \Phi$. Then, $(\mathbf{K}\phi \leftrightarrow K\phi) \in T$ if $p(\mathbf{K}) = 0$, and $(\mathbf{K}\phi \leftrightarrow \neg K\phi) \in T$ if $p(\mathbf{K}) = 1$. Then, it follows directly from Lemma 2 (see in Appendix A) that $(\mathbf{K}\phi \leftrightarrow \mathbf{K}'\phi) \in T$ whenever $p(\mathbf{K}) = p(\mathbf{K}')$. Finally, it follows from K_I that $\big(K(\mathbf{K}\phi) \leftrightarrow K(\mathbf{K}'\phi)\big) \in \omega$ for all $\omega \in \Omega_U$. □

Now we consider a state $\omega \in \Omega_U$ and a primitive proposition ϕ such that the agent wrongly assumes consciousness of all primitives at ω. We analyze the cases that can arise, representing different types of failures of consciousness of primitives. Cases 1 and 2 below correspond to failures of positive introspection ($K\phi$ while $\neg KK\phi$), whereas cases 3 and 4 consider failures of negative introspection ($\neg K\phi$ while $\neg K\neg K\phi$).

Case 1: Delusion on negative knowledge Assume $K\phi$, $\neg KK\phi$, and $K\neg K\phi$ all belong to ω.

In this case, the agent exhibits delusion about his own knowledge, as he thinks that he doesn't know ϕ, while ϕ is known. He is in the situation of a child who knows how to bike and who hasn't realized it, or a Mr Jourdain who knows how to talk in prose but who hasn't discovered it yet. We say that the agent then exhibits delusion on negative knowledge.

Consider any sequence $\mathbf{K}$ of knowledge operators and negations. Since $p(K\neg K) = 1$ and $p(K) = 0$, it follows from Lemma 1 that $K(\mathbf{K}\phi) \in \omega$ if and only if $p(\mathbf{K}) = 1$.

Case 2: Unawareness of knowledge Assume $K\phi$, $\neg KK\phi$, and $\neg K\neg K\phi$ are in ω.

Taking $\mathbf{K}$ to be any sequence of knowledge operators, we then have $\neg K(\mathbf{K}\phi) \in \omega$. The agent is then unaware of his knowledge of ϕ. Note that he is not unaware of ϕ since ϕ is known.

Case 3: Delusion on positive knowledge Assume $\neg K\phi$, $\neg K\neg K\phi$, and $KK\phi$ are in ω.

Again here, the agent exhibits delusion about his own knowledge, as he thinks that he knows ϕ, while ϕ is not known. This is a situation similar to that of a traveller who is sure that he knows how to get from point A to point B, before getting lost and realizing the absence of this knowledge. We say in this case that the agent then exhibits delusion on negative knowledge.

Again using lemma 1, we obtain $K(\mathbf{K}\phi) \in \omega$ if and only if $p(\mathbf{K}) = 0$.

Case 4: Unawareness Finally, consider ω containing $\neg K\phi$, $\neg K\neg K\phi$, and $\neg KK\phi$.

As in case 2, we have $\neg K(\mathbf{K}\phi) \in \omega$ for any sequence $\mathbf{K}$. Note that, contrary to case 2, we also have $\neg K\phi$, implying that the agent is unaware of ϕ.

The next theorem summarizes the different cases.

Theorem 3. *Assume the reasoning agent wrongly assumes consciousness of all primitive propositions at ω, then there exists a primitive proposition ϕ such that one of the following mutually exclusive cases occurs:*

1. *The agent has delusion on his negative knowledge of ϕ: $K\neg K\phi$ and $K\phi$ belong to ω.*
2. *The agent has delusion on his positive knowledge of ϕ: $KK\phi$ and $\neg K\phi$ belong to ω.*
3. *The agent is unaware of $K\phi$ at ω, while being unaware of ϕ or not.*

While it is important to note that the state space Ω_U allows for the possibility of unawareness, contrary to "standard" state space models, it is also remarkable that this unawareness cannot hold in any arbitrary fashion. If the agent is unaware of any proposition, then he must be unaware of either a primitive proposition, or of his knowledge of such a primitive proposition.

Unawareness of a primitive proposition arises from a failure of negative introspection on this proposition: The agent does not know it, but fails to recognize that he does not know it. On the other hand, failure of positive introspection, in which case the agent knows a proposition without realizing that he knows this proposition, can give rise to unawareness of the knowledge of this proposition, but not of awareness of the proposition itself.

When the agent wrongly assumes consciousness of primitive propositions, his wrong assumptions on the structure of his knowledge may lead him to delusion. Surprisingly enough, this delusion only occurs either on primitives, ($K\phi \in \omega$ and $\neg\phi \in \omega$), or on knowledge of primitives. This last situation happens when the agent has either delusion on positive knowledge, or on negative knowledge. Note that delusion never occurs on higher order levels of knowledge: For any state ω in Ω_U, any proposition ϕ and any sequence $\mathbf{K}$ of length at least 2, $K(\mathbf{K}\phi) \in \omega$ implies $\mathbf{K}\phi \in \omega$.

5 Complexity of the state space

We first show that states in Ω_C, as well as the agent's knowledge in this state space, have an easy description.

Theorem 4. *In Ω_C, a state is determined by primitive propositions and knowledge of primitive propositions, and the agent's knowledge is determined by his knowledge of primitive propositions:*

1. *For every $\omega, \omega' \in \Omega_C$, if $\omega(\phi) = \omega'(\phi)$ and $\omega(K\phi) = \omega'(K\phi)$ for every $\phi \in \Phi_0$, then $\omega = \omega'$.*

2. *For every $\omega, \omega' \in \Omega_C$, if $\omega(K\phi) = \omega'(K\phi)$ for every $\phi \in \Phi_0$, then $\omega(K\phi) = \omega'(K\phi)$ for every $\phi \in \Phi$.*

The state space Ω_U is much richer than Ω_C. Still, a state of Ω_U can be described using a relatively low number of propositions. In particular, the state spaces are finite if the primitive propositions are derived from a finite family of "basic" primitive propositions, describing e.g., the fundamentals of the economy.

Following Aumann (1999), define the epistemic depth of a proposition ϕ is the number of nested knowledge operators found in this proposition. It is 0 for primitive propositions ϕ, the depth of $\neg\phi$ is the same as the depth of ϕ, the depth of $\phi_1 \vee \phi_2$ and $\phi_1 \wedge \phi_2$ is the maximum of the depths of ϕ_1 and ϕ_2, and the depth of $K\phi$ is equal to the depth of ϕ plus one. We let Φ_n denote the set of propositions of epistemic depth at most n.

Formally, we define Φ_n as the closure of the set $\{\phi, K\phi \mid \phi \in \Phi_{n-1}\}$ with respect to $\neg$, $\vee$ and $\wedge$.

Theorem 5. *In Ω_U, a state is determined by primitive propositions and knowledge of propositions of epistemic depth at most one, and the agent's knowledge is determined by his knowledge of propositions of epistemic depth at most one:*

1. *For every $\omega, \omega' \in \Omega_U$, if $\omega(\phi) = \omega'(\phi)$ for every $\phi \in \Phi_0$ and $\omega(K\phi) = \omega'(K\phi)$ for every $\phi \in \Phi_1$, then $\omega = \omega'$.*

2. *For every $\omega, \omega' \in \Omega_U$, if $\omega(K\phi) = \omega'(K\phi)$ for every $\phi \in \Phi_1$, then $\omega(K\phi) = \omega'(K\phi)$ for every $\phi \in \Phi$.*

Theorem 5 shows that, although allowing for rich possibilities in the description of the agent's knowledge, the state space Ω_U still remains tractable. The main reason is that, through a process of deductive reasoning, the agent is able to derive all higher order knowledge from knowledge of epistemic propositions of depth at most one.

6 Discussion

6.1 Relationship to the existing literature

The notion of unawareness has attracted the attention of numerous authors recently. The first paper to appear in the literature was that of Fagin and Halpern (1988), who provide an axiomatic

characterization of awareness. They introduce separate operators for explicit knowledge – which is equivalent to the standard notion of knowledge – and also implicit knowledge, which can be thought as the set of logical consequences of the explicitly known propositions. A proposition is explicitly known whenever the agent implicitly knows it and is aware of it.

Modica and Rustichini (1994) provided an explicit definition of unawareness – similar to ours: An agent is unaware of a proposition if he does not know it and does not know that he does not know it. The showed that if being unaware of a proposition implies that one is unaware of its negation too, then the agent's knowledge is modeled with S5 again, implying that the only way to model unawareness is by relaxing the inference rules. They provide such a model in their follow up paper (Modica and Rustihini, 1999). Halpern (2001) proved that the latter is a special case of Fagin and Halpern (1988).

In the context of semantic models Dekel et al. (1998) had already shown that there is no unawareness operator that can capture the notion of non-trivial unawareness, i.e., it is not possible to model unawareness in such a restrictive framework as the one of semantic models. Heifetz et al. (2006) suggest an alternative generalized semantic model that accommodates most desiderata considered so far in the literature, and therefore does not suffer from the impossibility result proven by Dekel et al. (1998).

Feinberg (2004, 2005) introduced unawareness in a game theoretic setting and discussed the implications of unawareness about the action space of a game to different equilibrium outcomes and solution concepts.

Appendix A: Proof of Theorem 1

Definition 1. *Let* $(\phi_1 \overset{T}{\rightarrow} \phi_2)$ *be a shorthand for the following statement:*

$$\textit{if } \phi_1 \in T \textit{ then } \phi_2 \in T.$$

Lemma 2. *Consider a reasoning agent. Then,*

(i) $(\phi_1 \rightarrow \phi_2) \overset{T}{\rightarrow} (\neg\phi_2 \rightarrow \neg\phi_1)$,

(ii) $\big((\phi_1 \rightarrow \phi_2) \wedge (\phi_2 \rightarrow \phi_3)\big) \overset{T}{\rightarrow} (\phi_1 \rightarrow \phi_3)$,

(iii) *if* $\phi_1 \overset{T}{\rightarrow} \phi_3$ *and* $\phi_2 \overset{T}{\rightarrow} \phi_4$, *then* $(\phi_1 \wedge \phi_2) \overset{T}{\rightarrow} (\phi_3 \wedge \phi_4)$.

Proof. (i) It follows directly from the definition of the implication.

(*ii*) Consider the following sequence of tautologies:

$$\begin{aligned}\big((\phi_1 \to \phi_2) \wedge (\phi_2 \to \phi_3)\big) &\stackrel{\mathrm{T}}{\to} \big((\neg\phi_1 \wedge \neg\phi_2) \vee (\neg\phi_1 \wedge \phi_3) \vee (\phi_2 \wedge \neg\phi_2) \vee (\phi_2 \wedge \phi_3)\big)\\ &\stackrel{\mathrm{T}}{\to} \big((\neg\phi_1 \wedge \neg\phi_2) \vee (\neg\phi_1 \wedge \phi_3) \vee (\phi_2 \wedge \phi_3)\big)\\ &\stackrel{\mathrm{T}}{\to} (\neg\phi_1 \vee \phi_3)\\ &\stackrel{\mathrm{T}}{\to} (\phi_1 \to \phi_3).\end{aligned}$$

(*iii*) The following relationships hold:

$$\begin{aligned}(\phi_1 \wedge \phi_2) &\stackrel{\mathrm{T}}{\to} \phi_1 \stackrel{\mathrm{T}}{\to} \phi_3,\\ (\phi_1 \wedge \phi_2) &\stackrel{\mathrm{T}}{\to} \phi_2 \stackrel{\mathrm{T}}{\to} \phi_4.\end{aligned}$$

That is, if $(\phi_1 \wedge \phi_2) \in T$ then $(\phi_3 \wedge \phi_4) \in T$. □

Lemma 3. *Consider a reasoning agent. Then, for every* $\phi \in \Phi$

(*i*) $C\phi \stackrel{T}{\to} CK\phi$, *and*

(*ii*) $C\phi \stackrel{T}{\to} C\neg K\phi$.

Proof. By definition $C\phi$ is an abbreviation for $(K\phi \to KK\phi) \wedge (\neg K\phi \to K\neg K\phi)$. It follows from A_1 that $C\phi \in T$ is equivalent to $(K\phi \to KK\phi) \in T$ and $(\neg K\phi \to K\neg K\phi) \in T$.

(*i*) It follows from R_I that

$$(K\phi \to KK\phi) \stackrel{\mathrm{T}}{\to} (KK\phi \to KKK\phi). \tag{1}$$

Furthermore,

$$\begin{aligned}(K\phi \to KK\phi) &\stackrel{\text{(by Lemma 2)}}{\stackrel{\mathrm{T}}{\to}} (\neg KK\phi \to \neg K\phi)\\ &\stackrel{\text{(by } C\phi\in T\text{)}}{\stackrel{\mathrm{T}}{\to}} (\neg KK\phi \to \neg K\phi) \wedge (\neg K\phi \to K\neg K\phi)\\ &\stackrel{\text{(by Lemma 2)}}{\stackrel{\mathrm{T}}{\to}} (\neg KK\phi \to K\neg K\phi)\\ &\stackrel{\text{(by } C\phi\in T \text{ and Lemma 2)}}{\stackrel{\mathrm{T}}{\to}} (\neg KK\phi \to K\neg K\phi) \wedge (\neg K\phi \to \neg KK\phi)\\ &\stackrel{\text{(by } R_I\text{)}}{\stackrel{\mathrm{T}}{\to}} (\neg KK\phi \to K\neg K\phi) \wedge (K\neg K\phi \to K\neg KK\phi)\\ &\stackrel{\text{(by Lemma 2)}}{\stackrel{\mathrm{T}}{\to}} (\neg KK\phi \to K\neg KK\phi).\end{aligned} \tag{2}$$

Finally, it follows that

$$\begin{aligned}
C\phi &\overset{\text{(by definition)}}{\overset{\mathrm{T}}{\to}} (K\phi \to KK\phi) \wedge (\neg K\phi \to K\neg K\phi) \\
&\overset{\text{(by } R_1\text{)}}{\overset{\mathrm{T}}{\to}} (K\phi \to KK\phi) \\
&\overset{\text{(by (1) and (2))}}{\overset{\mathrm{T}}{\to}} (KK\phi \to KKK\phi) \wedge (\neg KK\phi \to K\neg KK\phi) \\
&\overset{\text{(by definition)}}{\overset{\mathrm{T}}{\to}} CK\phi,
\end{aligned}$$

which completes the proof.

(ii) The proof of $C\neg K\phi \in T$ is very similar to (i): It follows from R_I that

$$(\neg K\phi \to K\neg K\phi) \overset{\mathrm{T}}{\to} (K\neg K\phi \to KK\neg K\phi). \tag{3}$$

Furthermore,

$$\begin{aligned}
(\neg K\phi \to K\neg K\phi) &\overset{\text{(by Lemma 2)}}{\overset{\mathrm{T}}{\to}} (\neg K\neg K\phi \to K\phi) \\
&\overset{\text{(by } C\phi\in T\text{)}}{\overset{\mathrm{T}}{\to}} (\neg K\neg K\phi \to K\phi) \wedge (K\phi \to KK\phi) \\
&\overset{\text{(by Lemma 2)}}{\overset{\mathrm{T}}{\to}} (\neg K\neg K\phi \to KK\phi) \\
&\overset{\text{(by } (i)\text{)}}{\overset{\mathrm{T}}{\to}} (\neg K\neg K\phi \to KK\phi) \wedge (KK\phi \to KKK\phi) \\
&\overset{\text{(by Lemma 2)}}{\overset{\mathrm{T}}{\to}} (\neg K\neg K\phi \to KKK\phi) \\
&\overset{\text{(by } A'_3\text{)}}{\overset{\mathrm{T}}{\to}} (\neg K\neg K\phi \to KKK\phi) \wedge (KK\phi \to \neg K\neg K\phi) \\
&\overset{\text{(by } R_I\text{)}}{\overset{\mathrm{T}}{\to}} (\neg K\neg K\phi \to KKK\phi) \wedge (KKK\phi \to K\neg K\neg K\phi) \\
&\overset{\text{(by Lemma 2)}}{\overset{\mathrm{T}}{\to}} (\neg K\neg K\phi \to K\neg K\neg K\phi).
\end{aligned} \tag{4}$$

Finally, it follows that

$$\begin{aligned}
C\phi &\overset{\text{(by definition)}}{\overset{\mathrm{T}}{\to}} (K\phi \to KK\phi) \wedge (\neg K\phi \to K\neg K\phi) \\
&\overset{\text{(by } R_1\text{)}}{\overset{\mathrm{T}}{\to}} (\neg K\phi \to K\neg K\phi) \\
&\overset{\text{(by (3) and (4))}}{\overset{\mathrm{T}}{\to}} (K\neg K\phi \to KK\neg K\phi) \wedge (\neg K\neg K\phi \to K\neg K\neg K\phi) \\
&\overset{\text{(by definition)}}{\overset{\mathrm{T}}{\to}} C\neg K\phi,
\end{aligned}$$

which completes the proof. □

Lemma 4. *Consider a reasoning agent. If $C\phi_1 \in T$ and $C\phi_2 \in T$, then $C(\phi_1 \wedge \phi_2) \in T$, for all $\phi_1, \phi_2 \in \Phi$.*

Proof. It follows from A_1' that $\big(K(\phi_1 \wedge \phi_2) \rightarrow (K\phi_1 \wedge K\phi_2)\big) \in T$. Then,

$$\begin{aligned}
& \big(K(\phi_1 \wedge \phi_2) \rightarrow (K\phi_1 \wedge K\phi_2)\big) \\
\overset{\text{(by } C\phi \in T\text{)}}{\overset{\mathrm{T}}{\rightarrow}} \quad & \big(K(\phi_1 \wedge \phi_2) \rightarrow (K\phi_1 \wedge K\phi_2)\big) \wedge \big((K\phi_1 \wedge K\phi_2) \rightarrow (KK\phi_1 \wedge KK\phi_2)\big) \\
\overset{\text{(by Lemma 2)}}{\overset{\mathrm{T}}{\rightarrow}} \quad & \big(K(\phi_1 \wedge \phi_2) \rightarrow (KK\phi_1 \wedge KK\phi_2)\big) \\
\overset{\text{(by } A_1'\text{)}}{\overset{\mathrm{T}}{\rightarrow}} \quad & \big(K(\phi_1 \wedge \phi_2) \rightarrow KK(\phi_1 \wedge \phi_2)\big). \qquad (5)
\end{aligned}$$

It follows from A_0' and A_1' that $\big(\neg K(\phi_1 \wedge \phi_2) \rightarrow (\neg K\phi_1 \vee \neg K\phi_2)\big) \in T$. Then, we show – similarly to above – that

$$\begin{aligned}
& \big(\neg K(\phi_1 \wedge \phi_2) \rightarrow (\neg K\phi_1 \vee \neg K\phi_2)\big) \\
\overset{\mathrm{T}}{\rightarrow} \quad & \big(\neg K(\phi_1 \wedge \phi_2) \rightarrow K\neg K(\phi_1 \wedge \phi_2)\big),
\end{aligned}$$

which completes the proof. □

Lemma 5. *Consider a reasoning agent such that $C\phi_1 \in T$ and $C\phi_2 \in T$. Then,*

(i) $C(K\phi_1 \vee \phi_2) \in T$, and

(ii) $C(\neg K\phi_1 \vee \phi_2) \in T$.

Proof. First we show that

$$\big(K(K\phi_1 \vee \phi_2) \leftrightarrow (K\phi_1 \vee K\phi_2)\big) \in T. \qquad (6)$$

It follows from $C\phi_1 \in T$ that $\big((K\phi_1 \vee K\phi_2) \rightarrow (KK\phi_1 \vee K\phi_2)\big) \in T$. Thus, it follows from A_2' that

$$\big((K\phi_1 \vee K\phi_2) \rightarrow K(K\phi_1 \vee \phi_2)\big) \in T.$$

For the converse, it follows by definition that $\big(K(K\phi_1 \vee \phi_2) \rightarrow K(\neg K\phi_1 \rightarrow \phi_2)\big) \in T$. Then,

$$\begin{aligned}
\big(K(K\phi_1 \vee \phi_2) \rightarrow K(\neg K\phi_1 \rightarrow \phi_2)\big) &\overset{\text{(by definition)}}{\overset{\mathrm{T}}{\rightarrow}} \big(\neg K(K\phi_1 \vee \phi_2) \vee K(\neg K\phi_1 \rightarrow \phi_2) \\
&\overset{\text{(by } R_I\text{)}}{\overset{\mathrm{T}}{\rightarrow}} \big(\neg K(K\phi_1 \vee \phi_2) \vee (K\neg K\phi_1 \rightarrow K\phi_2)\big) \\
&\overset{\text{(by definition)}}{\overset{\mathrm{T}}{\rightarrow}} \big(K(K\phi_1 \vee \phi_2) \rightarrow (\neg K\neg K\phi_1 \vee K\phi_2)\big) \\
&\overset{\text{(by } C\phi\in T\text{)}}{\overset{\mathrm{T}}{\rightarrow}} \big(K(K\phi_1 \vee \phi_2) \rightarrow (K\phi_1 \vee K\phi_2)\big).
\end{aligned}$$

(i) Consider the following sequence of equivalences:

$$\begin{aligned}
&\big(K(K\phi_1 \vee \phi_2) \rightarrow (K\phi_1 \vee K\phi_2)\big) \\
&\overset{\text{(by } C\phi\in T\text{)}}{\overset{\mathrm{T}}{\rightarrow}} \big(K(K\phi_1 \vee \phi_2) \rightarrow (K\phi_1 \vee K\phi_2)\big) \wedge \big((K\phi_1 \vee K\phi_2) \rightarrow (KKK\phi_1 \vee KK\phi_2)\big) \\
&\overset{\text{(by Lemma 2)}}{\overset{\mathrm{T}}{\rightarrow}} \big(K(K\phi_1 \vee \phi_2) \rightarrow (KKK\phi_1 \vee KK\phi_2)\big) \\
&\overset{\text{(by } C\phi\in T\text{)}}{\overset{\mathrm{T}}{\rightarrow}} \big(K(K\phi_1 \vee \phi_2) \rightarrow KK(K\phi_1 \vee \phi_2)\big). \qquad (7)
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\big(\neg K(K\phi_1 \vee \phi_2) \rightarrow \neg(K\phi_1 \vee K\phi_2)\big) \\
&\overset{\text{(by } A'_0\text{)}}{\overset{\mathrm{T}}{\rightarrow}} \big(\neg K(K\phi_1 \vee \phi_2) \rightarrow (\neg K\phi_1 \wedge \neg K\phi_2)\big) \\
&\overset{\text{(by } C\phi\in T\text{)}}{\overset{\mathrm{T}}{\rightarrow}} \big(\neg K(K\phi_1 \vee \phi_2) \rightarrow (K\neg K\phi_1 \wedge K\neg K\phi_2)\big) \\
&\overset{\text{(by } A'_0 \text{ and } A'_2\text{)}}{\overset{\mathrm{T}}{\rightarrow}} \big(\neg K(K\phi_1 \vee \phi_2) \rightarrow K\neg(K\phi_1 \vee K\phi_2)\big) \\
&\overset{\text{(by (6))}}{\overset{\mathrm{T}}{\rightarrow}} \big(\neg K(K\phi_1 \vee \phi_2) \rightarrow K\neg(K\phi_1 \vee K\phi_2)\big) \wedge \big(\neg(K\phi_1 \vee K\phi_2) \rightarrow \neg K(K\phi_1 \vee \phi_2)\big) \\
&\overset{\text{(by } R_I\text{)}}{\overset{\mathrm{T}}{\rightarrow}} \big(\neg K(K\phi_1 \vee \phi_2) \rightarrow K\neg(K\phi_1 \vee K\phi_2)\big) \wedge \big(K\neg(K\phi_1 \vee K\phi_2) \rightarrow K\neg K(K\phi_1 \vee \phi_2)\big) \\
&\overset{\text{(by Lemma 2)}}{\overset{\mathrm{T}}{\rightarrow}} \big(\neg K(K\phi_1 \vee \phi_2) \rightarrow K\neg K(K\phi_1 \vee \phi_2)\big). \qquad (8)
\end{aligned}$$

It follows from (7) and (8) that $\big(K(K\phi_1 \vee \phi_2) \rightarrow KK(K\phi_1 \vee \phi_2)\big) \wedge \big(\neg K(K\phi_1 \vee \phi_2) \rightarrow K\neg K(K\phi_1 \vee \phi_2)\big) \in T$, implying $C(K\phi_1 \vee \phi_2) \in T$.

(ii) It follows from $C\phi_1 \in T$ that $\big((\neg K\phi_1 \vee \phi_2) \leftrightarrow (K\neg K\phi_1 \vee \phi_2)\big) \in T$. Then, apply the steps of the proof of (i) for $(K\neg K\phi_1 \vee \phi_2)$ and the proof is completed. □

Proof of Theorem 1. Recall that we define Φ_n as the closure of the set $\{\phi, K\phi \mid \phi \in \Phi_{n-1}\}$ with respect to $\neg$, $\vee$ and $\wedge$. It is straightforward verifying that $\Phi_\infty := \bigcup_{n\geq 0} \Phi_n$ is such that $\Phi_\infty = \Phi$.

Thus, we prove the theorem by induction: We show that if $C\phi \in T$ for all $\phi \in \Phi_n$, then $C\phi' \in T$ for all $\phi' \in \Phi_{n+1}$. This follows directly from Lemmas 3, 4 and 5. □

Appendix B: Proof of Theorems 3, 4 and 5

Proof of Theorem 3. Since the reasoning agent wrongly assumes consciousness at ω, there is some primitive ϕ such that $\neg C\phi \in \omega$ implying that

A. $(K\phi \wedge \neg KK\phi) \in \omega$, or

B. $(\neg K\phi \wedge \neg K\neg K\phi) \in \omega$.

If (A) occurs there are two possible subcases:

A.1. $K\neg KK\phi \in \omega$, in which case (by Lemma 1) it follows $K\neg K\phi \in \omega$, which corresponds to Case (1).

A.2. $\neg K\neg KK\phi \in \omega$, in which case the agent is unaware of $K\phi$ at ω, which corresponds to Case (3). To see this, suppose that there is some $\mathbf{K}\phi$ such that $K(\mathbf{K}\phi) \in \omega$. If $p(\mathbf{K}) = 0$ then it follows from Lemma 1 that $KK\phi \in \omega$ which contradicts $\neg KK\phi \in \omega$, whereas if $p(\mathbf{K}) = 1$ then it follows from Lemma 1 that $K\neg KK\phi \in \omega$ which contradicts $\neg K\neg KK\phi \in \omega$. Note that in this case the agent is unaware of $K\phi$ without being unaware of ϕ.

If (B) occurs there are two possible subcases:

A.1. $K\neg K\neg K\phi \in \omega$, in which case (by Lemma 1) it follows $KK\phi \in \omega$, which corresponds to Case (2).

A.2. $\neg K\neg K\neg K\phi \in \omega$, in which case the agent is unaware of $K\phi$ at ω, which corresponds to Case (3). To see this, suppose that there is some $\mathbf{K}\phi$ such that $K(\mathbf{K}\phi) \in \omega$. If $p(\mathbf{K}) = 0$ then it follows from Lemma 1 that $K\neg K\neg K\phi \in \omega$ which contradicts $\neg K\neg K\neg K\phi \in \omega$, whereas if $p(\mathbf{K}) = 1$ then it follows from Lemma 1 that $K\neg K\phi \in \omega$ which contradicts $\neg K\neg K\phi \in \omega$. Note that in this case the agent is unaware of $K\phi$ and of ϕ at the same time.

The previous cases complete the proof. □

Proof of Theorem 4. Point (1) is already known in the context of Modal Logic, see Halpern (1995). Point (2) is new. We include a proof of both parts for the sake of completeness.

1. Take some $\omega, \omega', \in \Omega_C$ and two arbitrary propositions ϕ_1, ϕ_2 such that $\omega(\phi) = \omega'(\phi)$ and $\omega(K\phi) = \omega'(K\phi)$. Consider the following cases:

A. Let $K\phi \in \omega$, and take the following equivalent statements for :

$$\begin{aligned} K\phi \in \omega &\overset{\text{(by Theorem 1)}}{\Longrightarrow} KK\phi \in \omega \\ &\overset{\text{(by } A_6\text{)}}{\Longrightarrow} \neg K\neg K\phi \in \omega \\ &\overset{\text{(by Theorem 1)}}{\Longrightarrow} K\phi \in \omega, \end{aligned}$$

implying that $K\phi \in \omega$ if and only if $KK\phi \in \omega$. The same holds for ω'. Therefore, if $\omega(\phi) = \omega'(\phi)$ and $\omega(K\phi) = \omega'(K\phi)$, then $\omega(K\phi) = \omega'(K\phi)$ and $\omega(KK\phi) = \omega'(KK\phi)$.

B. Let $K\phi_1 \in \omega$ and $K\phi_2 \in \omega$. Likewise,

$$(\phi_1 \wedge \phi_2) \in \omega \overset{\text{(by } A_2\text{)}}{\Longleftrightarrow} \phi_1 \in \omega \text{ and } \phi_2 \in \omega,$$

and

$$K(\phi_1 \wedge \phi_2) \in \omega \overset{\text{(by } A_4\text{)}}{\Longleftrightarrow} K\phi_1 \in \omega \text{ and } K\phi_2 \in \omega,$$

implying that if $\omega(\phi_1) = \omega'(\phi_1)$ and $\omega(K\phi_1) = \omega'(K\phi_1)$, and also $\omega(\phi_2) = \omega'(\phi_2)$ and $\omega(K\phi_2) = \omega'(K\phi_2)$, then $\omega(\phi_1 \wedge \phi_2) = \omega'(\phi_1 \wedge \phi_2)$ and $\omega\big(K(\phi_1 \wedge \phi_2)\big) = \omega'\big(K(\phi_1 \wedge \phi_2)\big)$.

C. Likewise,

$$(\phi_1 \vee \phi_2) \in \omega \overset{\text{(by } A_3\text{)}}{\Longleftrightarrow} \phi_1 \in \omega \text{ or } \phi_2 \in \omega.$$

Consider now the following subcases:

C.1. Let $K\phi_1 \in \omega$ or $K\phi_2 \in \omega$, yielding $K(K\phi_1 \vee \phi_2) \in \omega$.

C.1. Let $\neg K\phi_1 \in \omega$ and $\neg K\phi_2 \in \omega$, and suppose $K(K\phi_1 \vee \phi_2) \in \omega$. Then, obtain the following

equivalences:

$$
\begin{aligned}
K(K\phi \vee \phi_2) \in \omega \quad &\overset{\text{(by definition)}}{\Longrightarrow} \quad K(\neg K\phi_1 \rightarrow \phi_2) \in \omega \\
&\overset{\text{(by } K_I\text{)}}{\Longrightarrow} \quad (K\neg K\phi_1 \rightarrow K\phi_2) \in \omega \\
&\overset{\text{(by definition)}}{\Longrightarrow} \quad (\neg K\neg K\phi_1 \vee K\phi_2) \in \omega \\
&\overset{\text{(by Theorem 1)}}{\Longrightarrow} \quad (K\phi_1 \vee K\phi_2) \in \omega \\
&\overset{\text{(by } K_I\text{)}}{\Longrightarrow} \quad K\phi_1 \in \omega \text{ or } K\phi_2 \in \omega,
\end{aligned}
$$

which is a contradiction. Therefore, $K(K\phi \vee \phi_2) \in \omega$ if and only if $K\phi_1 \in \omega$ or $K\phi_2 \in \omega$. The same holds for ω', implying that if $\omega(\phi_1) = \omega'(\phi_1)$ and $\omega(K\phi_1) = \omega'(K\phi_1)$, and also $\omega(\phi_2) = \omega'(\phi_2)$ and $\omega(K\phi_2) = \omega'(K\phi_2)$, then $\omega(K\phi_1 \vee \phi_2) = \omega'(K\phi_1 \vee \phi_2)$ and $\omega\big(K(K\phi_1 \vee \phi_2)\big) = \omega'\big(K(K\phi_1 \vee \phi_2)\big)$.

It follows from $A - C$ that if $\omega, \omega', \in \Omega_C$ are such that $\omega(\phi) = \omega'(\phi)$ and $\omega(K\phi) = \omega'(K\phi)$ for all $\phi \in \Phi_0$ then they are also such that $\omega(\phi') = \omega'(\phi')$ and $\omega(K\phi') = \omega'(K\phi')$ for all ϕ' in the closure of $\{\phi, K\phi \mid \phi \in \Phi_0\}$, which by definition is Φ_1. Then, by induction the previous condition holds for all $\phi \in \Phi$, which completes the proof.

2. The only difference to the previous statement is the truth value of the primitive propositions. Thus, the proof is identical to the first part of the proof of (1). □

Proof of Theorem 5. Consider two arbitrary $\omega, \omega' \in \Omega_U$, and let $\omega(K\phi) = \omega'(K\phi)$ for all $\phi \in \Phi_1$. It follows from Lemma 1 that $(\mathbf{K}\phi \leftrightarrow \mathbf{K}'\phi) \in T$ for all $\mathbf{K}$ and $\mathbf{K}'$ that share parity. Then, $\big(K(\mathbf{K}\phi) \leftrightarrow K(\mathbf{K}'\phi)\big) \in \omega$ for all $\omega \in \Omega_U$. The rest of the proof follows by induction, and is identical to the one of Theorem 4 after having substituted Φ_n with Φ_{n+1}. □