Theory Dec. (2012) 73:513–523
DOI 10.1007/s11238-011-9284-1

Reasoning-based introspection

Olivier Gossner · Elias Tsakas

Published online: 26 October 2011

Abstract We show that if an agent reasons according to standard inference rules, the truth and introspection axioms extend from the set of non-epistemic propositions to the whole set of propositions. This implies that the usual axiomatization of partitional possibility correspondences is redundant, and provides a justification for truth and introspection that is partly based on reasoning.

Keywords Knowledge · Introspection · Truth axiom · Partitional information structures · Epistemic game theory

JEL Classification D80 · D83 · D89

1 Introduction

The information of an agent who processes information rationally is commonly represented by a partition over a state space. This representation is fundamental in many areas of Game Theory, such as modeling of games with incomplete information (Harsanyi 1967–1968), the study common knowledge (Aumann 1976), and epistemic

O. Gossner
Paris School of Economics, Paris, France

O. Gossner
London School of Economics and Political Science, London, UK
e-mail: ogossner@ens.fr

E. Tsakas (✉)
Maastricht University, Maastricht, The Netherlands
e-mail: e.tsakas@maastrichtuniversity.nl

foundations of solution concepts such as correlated equilibrium (Aumann 1987) and Nash equilibrium (Aumann and Brandenburger 1995).

The usual axiomatization for partitional state spaces is that of an agent whose knowledge satisfies truth and introspection (Fagin et al. 1995; Samet 1990; Aumann 1999). The truth axiom says that any proposition known to the agent is true. According to the introspection axioms, the agent knows both what he knows and what he does not know.

It is generally admitted that the agent's capacity to draw logical inferences is part of the properties that define the agent's rationality (e.g., see Geanakoplos 1989). However, it is not clear why a rational agent should also be endowed with introspective abilities. The objective of this note is to investigate to what extent truth and introspection can be explained by a deductive process on the part of the agent.

We consider an agent who observes natural facts about the surrounding world. These natural facts are those described by non-epistemic—else called Boolean—propositions and correspond to sentences that do not involve the agent's own knowledge. The agent also observes his own knowledge about natural facts, i.e., he knows what he knows and what he does not know about natural facts.

We show that the truth and introspection axioms are then satisfied for every proposition, whether epistemic or not. Since we only assume truth and introspection for non-epistemic propositions, our result provides a justification for these axioms for all other propositions that is based on deductive reasoning.

We express our results in terms of axiomatizations of knowledge. Formally, we show that the knowledge induced by the standard axiomatization of syntactic knowledge, $S5$, remains unchanged if truth and introspection are assumed for non-epistemic propositions only. In this sense, our result shows that the usual axiomatization $S5$ is redundant, as it can be replaced by a strict subset of axioms. In case there is a finite number of non-epistemic propositions, our axiomatization assumes truth and introspection for a finite set of propositions, instead of a countable set in the usual axiomatization of $S5$.

The axiomatic model and main result are presented in Sect. 2. We present a discussion in Sect. 3: In Sect. 3.1, we recall the connection between $S5$ and partitional models, in Sect. 3.2 we discuss the tightness of the weaker axiomatization presented, in Sect. 3.3 we explain why our result is not a consequence of the well-known equivalence in $S5$ between every proposition and a proposition of epistemic depth at most one.

2 Model and main result

We recall the standard syntactic model of knowledge (Chellas 1980; Fagin et al. 1995). Let Φ be the alphabet of the agent's language, called the set of atomic propositions. These atomic propositions express basic facts about nature such as "it is raining in New York", or "the cat is mortal". The set of non-epistemic (else called Boolean) propositions $\mathcal{L}^0(\Phi)$ is the closure of Φ with respect to the standard connectives of negation, $\neg$, and conjunction, $\wedge$.

The knowledge operator is denoted by K, and $K\phi$ stands for "the agent knows ϕ". The set of all propositions $\mathcal{L}(\Phi)$, with generic elements ϕ, ψ, is the closure of Φ

with respect to $\neg$, $\wedge$ and K. The propositions $\phi \vee \psi$, $\phi \rightarrow \psi$ and $\phi \leftrightarrow \psi$ stand for $\neg(\neg\phi \wedge \neg\psi)$, $\neg\phi \vee \psi$, and $(\phi \rightarrow \psi) \wedge (\psi \rightarrow \phi)$ respectively.

We recall the standard Modal Logic (ML) system $S5$, consisting of the following axioms and inference rules:

A_1. All tautologies of propositional calculus
A_2. $\big(K\phi \wedge K(\phi \rightarrow \psi)\big) \rightarrow K\psi$ (Axiom of distribution)
A_3. $K\phi \rightarrow \phi$ (Truth axiom)
A_4. $K\phi \rightarrow KK\phi$ (Positive introspection)
A_5. $\neg K\phi \rightarrow K\neg K\phi$ (Negative introspection)
R_1. From ϕ and $(\phi \rightarrow \psi)$ infer ψ (Modus Ponens)
R_2. From ϕ infer $K\phi$ (Rule of necessitation)

The first axiom, A_1, refers to propositions such as $(\phi \rightarrow \psi) \leftrightarrow (\neg\psi \rightarrow \neg\phi)$, which are always logically true; A_2 says that if the agent knows that ϕ implies ψ, and knows ϕ, then he necessarily knows ψ; the truth axiom says that the agent cannot wrongly know a proposition; positive introspection states that the agent knows what he knows, whereas negative introspection says that the agent knows what he does not know. In the literature, an axiom typically holds for every $\phi \in \mathcal{L}(\Phi)$, e.g., A_3 is commonly understood as the set of propositions $\{K\phi \rightarrow \phi \mid \phi \in \mathcal{L}(\Phi)\}$. Throughout the paper, we study the implications of explicitly restricting an axiom to hold for some propositions only, e.g., if we say that A_3 holds for all propositions in $\Psi \subseteq \mathcal{L}(\Phi)$, we have in mind a system where the propositions in $\{K\phi \rightarrow \phi \mid \phi \in \Psi\}$ are axioms, whereas the ones in $\{K\phi \rightarrow \phi \mid \phi \in \mathcal{L}(\Phi) \setminus \Psi\}$ are not.

The propositions that can be proven from the axioms, or from other propositions that have already been proven, are called theorems. Formally, the set of theorems is the closure of the axioms with respect to the inference rules. We say that two ML systems are equivalent whenever they have the same set of theorems. A system is redundant if there is an equivalent system with a strictly smaller (with respect to inclusion) set of axioms, and the same inference rules.

Let T_5 denote the set of theorems in $S5$. We show that $S5$ is redundant, in that the same set of theorems, T_5, is obtained if Truth and Introspection are axioms only for the non-epistemic propositions. Formally, let $S5^0$ be the system consisting of the axioms:

A_1–A_2 for all propositions in $\mathcal{L}(\Phi)$, and
A_3–A_5 for all propositions in $\mathcal{L}^0(\Phi)$,

together with the usual inference rules R_1 and R_2 for every proposition. We let T_5^0 denote the set of theorems in $S5^0$, and we show that $S5$ and $S5^0$ are equivalent. Both in $S5$ and in $S5^0$, R_1 and R_2 apply to every proposition.

Main Theorem $T_5^0 = T_5$.

2.1 Proof of the Main Theorem

The general strategy of the proof is to show that (i) the set of propositions that satisfy truth and introspection is closed under a number of operations, and (ii) these operations are sufficient to generate the whole set of propositions starting from the non-epistemic propositions only. It is relatively straightforward to show that A_3–A_5 for ϕ and $\neg\phi$ imply $A3$–$A5$ for $K\phi$ and $\neg K\phi$ (Lemma 2), and that A_3–A_5 for ϕ and ψ imply A_3–A_5 for $\phi \wedge \psi$ (Lemma 3). However, it is not true in general that the set of propositions that satisfy A_3–A_5 is closed under $\vee$. Instead, we show that if ϕ, ψ satisfy truth and introspection, so do $K\phi \vee \psi$ and $\neg K\phi \vee \psi$ (Lemma 5).

We divide the proof of the Main Theorem into a series of lemmata.

Definition 1 For some $\Psi \subseteq \mathcal{L}(\Phi)$, let $S5(\Psi)$ denote the ML system consisting of

- the axioms A_1–A_2 for all propositions in $\mathcal{L}(\Phi)$,
- the axioms A_3–A_5 for all propositions in Ψ, and
- the inference rules R_1–R_2 for all propositions in $\mathcal{L}(\Phi)$.

and we let $T_5(\Psi)$ denote the set of Theorems in $S5(\Psi)$.

Note that for $\Psi \subseteq \Psi' \subseteq \mathcal{L}(\Phi)$, $T_5(\Psi) \subseteq T_5(\Psi')$.

Remark 1 In the proofs, we make repeated use of some well-known facts such as (i) $(K\phi \wedge K\psi) \leftrightarrow K(\phi \wedge \psi)$ is in $T_5(\emptyset)$, and for every $\Psi \subseteq \mathcal{L}(\Phi)$, (ii) if $\phi \to \psi$ is in $T_5(\Psi)$, then so is $\neg\psi \to \neg\phi$, (iii) if $\phi \to \phi'$ and $\phi' \to \phi''$ are in $T_5(\Psi)$, then so is $\phi \to \phi''$ (iv) if $K(\phi \to \psi)$ is in $T_5(\Psi)$, then so is $K\phi \to K\psi$.

In a series of lemmata, we prove Theorems in $S5(\Psi)$ for some subsets Ψ of $\mathcal{L}(\Phi)$. Unless Ψ is sufficiently rich, for instance if $\Psi = \emptyset$, it can be the case that in $T_5(\Psi)$, every proposition and its contrary is known to the agent. Our first lemma below shows that this cannot be the case if Ψ contains a contradiction, such as[1] $\psi \wedge \neg\psi$.

Lemma 1 *For every $\phi, \psi \in \mathcal{L}(\Phi)$, $(K\phi \to \neg K\neg\phi) \in T_5(\{\psi \wedge \neg\psi\})$.*

Proof By A_3 applied to $(\psi \wedge \neg\psi)$, and since $\neg(\psi \wedge \neg\psi)$ is a tautology of propositional calculus, $\neg K(\psi \wedge \neg\psi) \in T_5(\{\psi \wedge \neg\psi\})$. The proposition $(\phi \wedge \neg\phi) \to (\psi \wedge \neg\psi)$ is also a tautology of propositional calculus, hence belongs to $T_5(\{\psi \wedge \neg\psi\})$. Thus, both $K((\phi \wedge \neg\phi) \to (\psi \wedge \neg\psi))$ and $\neg K(\psi \wedge \neg\psi)$ are in $T_5(\{\psi \wedge \neg\psi\})$, which implies by A_2 that $\neg K(\phi \wedge \neg\phi) \in T_5(\{\psi \wedge \neg\psi\})$. This can be rewritten $(\neg K\phi \vee \neg K\neg\phi) \in T_5(\{\psi \wedge \neg\psi\})$, or $(K\phi \to \neg K\neg\phi) \in T_5(\{\psi \wedge \neg\psi\})$. □

Lemma 2 *For every $\phi, \psi \in \mathcal{L}(\Phi)$:*

1. *A_3–A_5 for $K\phi$ are in $T_5(\{\phi\})$,*
2. *A_3–A_5 for $\neg K\phi$ are in $T_5(\{\phi, \psi \wedge \neg\psi\})$.*

Proof We start with 1:

A_3 : A_3 for ϕ gives $(K\phi \to \phi) \in T_5(\{\phi\})$. By R_2, $K(K\phi \to \phi) \in T_5(\{\phi\})$. Thus, $(KK\phi \to K\phi) \in T_5(\{\phi\})$.

A_4 : By A_4 applied to ϕ, $(K\phi \to KK\phi) \in T_5(\{\phi\})$. By R_2, $K(K\phi \to KK\phi) \in T_5(\{\phi\})$, and $(KK\phi \to KKK\phi) \in T_5(\{\phi\})$.

[1] In fact, the result in the lemma holds under the weaker condition that the Truth axiom hold for some contradiction.

A_5 : By A_4 for ϕ, $(K\phi \to KK\phi) \in T_5(\{\phi\})$, hence $(\neg KK\phi \to \neg K\phi) \in T_5(\{\phi\})$. A_5 for ϕ gives $(\neg K\phi \to K\neg K\phi) \in T_5(\{\phi\})$, and $(\neg KK\phi \to K\neg K\phi) \in T_5(\{\phi\})$. By A_3 for $K\phi$ already proven above, $(KK\phi \to K\phi) \in T_5(\{\phi\})$, so that $(\neg K\phi \to \neg KK\phi) \in T_5(\{\phi\})$, and by R_2, $K(\neg K\phi \to \neg KK\phi) \in T_5(\{\phi\})$ and $(K\neg K\phi \to K\neg KK\phi) \in T_5(\{\phi\})$. Combining the two, we obtain $(\neg KK\phi \to K\neg KK\phi) \in T_5(\{\phi\})$.

Now, 2:

A_3 : A_4 applied to ϕ gives $(K\phi \to KK\phi) \in T_5(\{\phi\})$, and Lemma 1 applied to $K\phi$ shows $(KK\phi \to \neg K\neg K\phi) \in T_5(\{\psi \wedge \neg\psi\})$. Hence $(K\phi \to \neg K\neg K\phi) \in T_5(\{\phi, \psi \wedge \neg\psi\})$, and by contraposal $(K\neg K\phi \to \neg K\phi) \in T_5(\{\phi, \psi \wedge \neg\psi\})$.

A_4 : A_5 for ϕ gives $(\neg K\phi \to K\neg K\phi) \in T_5(\{\phi\})$. By R_2, $K(\neg K\phi \to K\neg K\phi) \in T_5(\{\phi\})$, hence $(K\neg K\phi \to KK\neg K\phi) \in T_5(\{\phi\})$.

A_5 : A_5 applied to ϕ gives $(\neg K\phi \to K\neg K\phi) \in T_5(\{\phi\})$, hence $(\neg K\neg K\phi \to K\phi) \in T_5(\{\phi\})$. Using A_4 for ϕ and $K\phi$ (proven above) we obtain $(K\phi \to KKK\phi) \in T_5(\{\phi\})$. From Lemma 1, $(KK\phi \to \neg K\neg K\phi) \in T_5(\{\psi \wedge \neg\psi\})$, hence by R_2 $K(KK\phi \to \neg K\neg K\phi) \in T_5(\{\psi \wedge \neg\psi\})$ and $KKK\phi \to K\neg K\neg K\phi \in T_5(\{\psi \wedge \neg\psi\})$. Combining the three, we obtain $\neg K\neg K\phi \to K\neg K\neg K\phi \in T_5(\{\phi, \psi \wedge \neg\psi\})$.

□

Lemma 3 *For every $\phi, \psi \in \mathcal{L}(\Phi)$, A_3–A_5 for $\phi \wedge \psi$ are in $T_5(\{\phi, \psi\})$.*

Proof A_3 : $\big(K(\phi \wedge \psi) \to (K\phi \wedge K\psi)\big) \in T_5(\{\emptyset\})$. By A_3 for ϕ, ψ, $K\phi \to \phi$ and $K\psi \to \psi$ are in $T_5(\{\phi, \psi\})$, and so is $\big(K\phi \wedge K\psi\big) \to (\phi \wedge \psi)$. Hence $\big(K(\phi \wedge \psi) \to (\phi \wedge \psi)\big) \in T_5(\{\phi, \psi\})$.

A_4 : $K(\phi \wedge \psi) \to (K\phi \wedge K\psi)$ and $KK\phi \wedge KK\psi \to K(K\phi \wedge K\psi)$ are in $\in T_5(\{\emptyset\})$. By A_4 applied to ϕ and ψ, $K\phi \to KK\phi$ and $K\psi \to KK\psi$, and also $K\phi \wedge K\psi \to KK\phi \wedge KK\psi$ are in $T_5(\{\phi, \psi\})$. It follows that $\big(K(\phi \wedge \psi) \to KK(\phi \wedge \psi)\big) \in T_5(\{\phi, \psi\})$.

A_5 : $(\neg K(\phi \wedge \psi) \to \neg K\phi \vee \neg K\psi) \in T_5(\{\emptyset\})$. By A_5 for ϕ and ψ, $\neg K\phi \to K\neg K\phi$, $\neg K\psi \to K\neg K\psi$, hence also $\neg K\phi \vee \neg K\psi \to K\neg K\phi \vee K\neg K\psi$ are in $T_5(\{\phi, \psi\})$. Since both $K\neg K\phi \to K(\neg K\phi \vee \neg K\psi)$ and $K\neg K\psi \to K(\neg K\phi \vee \neg K\psi)$ are in $T_5(\emptyset)$, so is $K\neg K\phi \vee K\neg K\psi \to K(\neg K\phi \vee \neg K\psi)$. Finally, $\neg K\phi \vee \neg K\psi \to \neg K(\phi \wedge \psi)$, and also, by R_2, $K(\neg K\phi \vee \neg K\psi) \to K\neg K(\phi \wedge \psi)$ are in $T_5(\emptyset)$. We conclude that $(\neg K(\phi \wedge \psi) \to K\neg K(\phi \wedge \psi)) \in T_5(\{\phi, \psi\})$.

□

Lemma 4 *For every $\phi, \psi \in \mathcal{L}(\Phi)$, $K(K\phi \vee \psi) \leftrightarrow (K\phi \vee K\psi)$ is in $T_5(\{\phi\})$.*

Proof $\leftarrow$: By A_4 for ϕ, $(K\phi \to KK\phi) \in T_5(\{\phi\})$, so $(K\phi \vee K\psi \to KK\phi \vee K\psi) \in T_5(\{\phi\})$. Since $KK\phi \to K(K\phi \vee \psi)$ and $K\psi \to K(K\phi \vee \psi)$ are in $T_5(\{\emptyset\})$, so is $KK\phi \vee K\psi \to K(K\phi \vee \psi)$. It follows that $((K\phi \vee K\psi) \to K(K\phi \vee \psi))) \in T_5(\{\phi\})$.

$\to$: $K(K\phi \vee \psi) \to K(\neg K\phi \to \psi)$, $K(\neg K\phi \to \psi) \to (K\neg K\phi \to K\psi)$, $(K\neg K\phi \to K\psi) \to (\neg K\neg K\phi \vee K\psi)$ are in $T_5(\{\emptyset\})$, hence so is

$K(K\phi \vee \psi) \rightarrow (\neg K\neg K\phi \vee K\psi)$. By A_5 applied to ϕ, $(\neg K\neg K\phi \rightarrow K\phi) \in T_5(\{\phi\})$, thus, $(K(K\phi \vee \psi) \rightarrow K\phi \vee K\psi) \in T_5(\{\phi\})$.

□

Lemma 5 *For every $\phi, \psi, \psi' \in \mathcal{L}(\Phi)$,*

1. *A_3–A_5 for $K\phi \vee \psi$ are in $T_5(\{\phi, \psi\})$,*
2. *A_3–A_5 for $\neg K\phi \vee \psi$ are in $T_5(\{\phi, \psi, \psi' \wedge \neg\psi'\})$.*

Proof First, for $K\phi \vee \psi$:

A_3 : From Lemma 4, $(K(K\phi \vee \psi) \rightarrow K\phi \vee K\psi) \in T_5(\{\phi\})$. By A_3 applied to ψ, $K\psi \rightarrow \psi \in T_5(\{\psi\})$. Thus, $(K(K\phi \vee \psi) \rightarrow K\phi \vee \psi) \in T_5(\{\phi, \psi\})$.

A_4 : From Lemma 4, $(K(K\phi \vee \psi) \rightarrow K\phi \vee K\psi) \in T_5(\{\phi\})$. By A_4 applied to ϕ, to $K\phi$ (Lemma 2) and to ψ, $K\phi \rightarrow KKK\phi$ and $K\psi \rightarrow KK\psi$ are in $T_5(\{\phi, \psi\})$. Hence, $(K\phi \vee K\psi) \rightarrow KKK\phi \vee KK\psi \in T_5(\{\phi, \psi\})$. Both $KKK\phi \rightarrow K(KK\phi \vee K\psi)$ and $KK\psi \rightarrow K(KK\phi \vee K\psi)$ are in $T_5(\{\emptyset\})$, hence so is $KKK\phi \vee KK\psi \rightarrow K(KK\phi \vee K\psi)$. Similarly, $KK\phi \vee K\psi \rightarrow K(K\phi \vee \psi) \in T_5(\{\phi\})$, so, from R_2, $K(KK\phi \vee K\psi) \rightarrow KK(K\phi \vee \psi) \in T_5(\{\phi\})$. We conclude that $(K(K\phi \vee \psi) \rightarrow KK(K\phi \vee \psi)) \in T_5(\{\phi, \psi\})$.

A_5 : From Lemma 4, $(K\phi \vee K\psi) \leftrightarrow K(K\phi \vee \psi)$, hence also $\neg K(K\phi \vee \psi) \leftrightarrow \neg(K\phi \vee K\psi)$ and $\neg K(K\phi \vee \psi) \leftrightarrow (\neg K\phi \wedge \neg K\psi)$ are in $T_5(\{\phi\})$. By A_5 for ϕ and ψ, $\neg K\phi \rightarrow K\neg K\phi \in T_5(\{\phi\})$, $\neg K\psi \rightarrow K\neg K\psi \in T_5(\{\psi\})$, hence $\neg K\phi \wedge \neg K\psi \rightarrow K(\neg K\phi \wedge \neg K\psi) \in T_5(\{\phi, \psi\})$. By R_2 applied to $\neg K\phi \wedge \neg K\psi \rightarrow \neg K(K\phi \vee \psi) \in T_5(\{\emptyset\})$, $K(\neg K\phi \wedge \neg K\psi) \rightarrow K\neg K(K\phi \vee \psi) \in T_5(\{\emptyset\})$. We conclude that $(\neg K(K\phi \vee \psi) \rightarrow K\neg K(K\phi \vee \psi)) \in T_5(\{\phi, \psi\})$.

Now, for $\neg K\phi \vee \psi$:

A_3 : By A_5 for ϕ, $\neg K\phi \vee \psi \rightarrow K\neg K\phi \vee \psi$, and, by R_2, also $K(\neg K\phi \vee \psi) \rightarrow K(K\neg K\phi \vee \psi)$ are in $T_5(\{\phi\})$. By Lemma 4, $(K(K\neg K\phi \vee \psi) \rightarrow K\neg K\phi \vee K\psi) \in T_5(\{\neg K\phi\})$, and by Lemma 2, $T_5(\{\neg K\phi\}) \subseteq T_5(\{\phi, \psi' \wedge \neg\psi'\})$. Since A_3 for $\neg K\phi$ and for ψ are in $T_5(\{\phi, \psi, \psi' \wedge \neg\psi'\})$, so is $K\neg K\phi \vee K\psi \rightarrow \neg K\phi \vee \psi$. We conclude that $K(\neg K\phi \vee \psi) \rightarrow \neg K\phi \vee \psi \in T_5(\{\phi, \psi, \psi' \wedge \neg\psi'\})$.

A_4 : As in the previous point, $(K(\neg K\phi \vee \psi) \rightarrow K\neg K\phi \vee K\psi) \in T_5(\{\phi\})$. Since A_4 for $\neg K\phi$ is in $T_5(\{\phi, \psi' \wedge \neg\psi'\})$, so is $K\neg K\phi \vee K\psi \rightarrow KK\neg K\phi \vee K\psi$. By Lemma 4, $KK\neg K\phi \vee K\psi \rightarrow K(K\neg K\phi \vee K\psi)$ is in $T_5(\{\psi\})$. Finally, $K\neg K\phi \rightarrow K(\neg K\phi \vee \psi)$ and $K\psi \rightarrow K(\neg K\phi \vee \psi)$ are in $T_5(\{\emptyset\})$, thus, so is $K\neg K\phi \vee K\psi \rightarrow K(\neg K\phi \vee \psi)$, and so is, by R_2, $K(K\neg K\phi \vee K\psi) \rightarrow KK(\neg K\phi \vee \psi)$. We conclude that $K(\neg K\phi \vee \psi) \rightarrow KK(\neg K\phi \vee \psi)$ is in $T_5(\{\phi, \psi, \psi' \wedge \neg\psi'\})$.

A_5 : From Lemma 4, $(K\neg K\phi \vee K\psi \rightarrow K(K\neg K\phi \vee \psi)) \in T_5(\{\psi\})$. By A_3 for $\neg K\phi$ (Lemma 2), $K\neg K\phi \vee \psi \rightarrow \neg K\phi \vee \psi$, hence by R_2 also $K(K\neg K\phi \vee \psi) \rightarrow K(\neg K\phi \vee \psi)$ are in $T_5(\{\phi, \psi' \wedge \neg\psi'\})$. We deduce that $K\neg K\phi \vee K\psi \rightarrow K(\neg K\phi \vee \psi)$ and its contraposive $\neg K(\neg K\phi \vee \psi) \rightarrow \neg K\neg K\phi \wedge \neg K\psi$ are in $T_5(\{\phi, \psi, \psi' \wedge \neg\psi'\})$. By A_5 for $\neg K\phi$ (Lemma 2) and for ψ, $(\neg K\neg K\phi \wedge \neg K\psi \rightarrow K(\neg K\neg K\phi \wedge \neg K\psi)) \in T_5(\{\phi, \psi, \psi' \wedge \neg\psi'\})$, and thus, $(\neg K\neg K\phi \wedge \neg K\psi \rightarrow K\neg(K\neg K\phi \vee K\psi)) \in T_5(\{\phi, \psi, \psi' \wedge \neg\psi'\})$.

By A_5 for $\phi, \neg K\phi \vee \psi \rightarrow K\neg K\phi \vee \psi$ in $T_5(\{\phi\})$, and by R_2 so is $K(\neg K\phi \vee \psi) \rightarrow K(K\neg K\phi \vee \psi)$. By Lemma 4, $(K(K\neg K\phi \vee \psi) \rightarrow K\neg K\phi \vee K\psi) \in T_5(\{\neg K\phi\}) \subseteq T_5(\{\phi, \psi' \wedge \neg\psi'\})$. We deduce that both $K(\neg K\phi \vee \psi) \rightarrow K\neg K\phi \vee K\psi$, its contraposive $\neg(K\neg K\phi \vee K\psi) \rightarrow \neg K(\neg K\phi \vee \psi)$, and, by R_2, $K\neg(K\neg K\phi \vee K\psi) \rightarrow K\neg K(\neg K\phi \vee \psi)$ are in $T_5(\{\phi, \psi, \psi' \wedge \neg\psi'\})$. We conclude that $\neg K(\neg K\phi \vee \psi) \rightarrow K\neg K(\neg K\phi \vee \psi)$ is in $T_5(\{\phi, \psi, \psi' \wedge \neg\psi'\})$.

□

We recursively define the set of propositions of epistemic depth at most n, by for $n \geq 1$, $\mathcal{L}^n(\Phi) = \mathcal{L}^0(\{\phi, K\phi \mid \phi \in \mathcal{L}^{n-1}(\Phi)\})$. Thus, $\mathcal{L}^0(\Phi)$ is the set of non-epistemic propositions, and $\mathcal{L}^n(\Phi)$ is the closure of $\{\phi, K\phi \mid \phi \in \Psi_{n-1}\}$ with respect to the logical connectives.

Lemma 6 *For every $n \geq 1$ and $\phi \in \mathcal{L}^n(\Phi)$, A_3–A_5 for ϕ are in $T_5(\mathcal{L}^{n-1}(\Phi))$.*

Proof Fix any $\psi' \in \Phi$, and note that $\psi' \wedge \neg\psi' \in \mathcal{L}^0(\Phi)$, so that A_3–A_5 for $\psi' \wedge \neg\psi'$ are in $T_5(\mathcal{L}^n(\Phi))$ for every n.

For $\phi \in \mathcal{L}^n(\Phi)$, there exist integers $k, l_1, \ldots l_k$ and, a family $(\phi_{i,j})_{i \leq k, j \leq l_k}$ of elements on $\mathcal{L}(\Phi)$ such that:

- $\phi \leftrightarrow (\phi_{1,1} \vee \cdots \vee \phi_{1,l_1}) \wedge \ldots \wedge (\phi_{k,1} \vee \cdots \vee \phi_{k,l_k})$ is a tautology of propositional calculus, hence in $T_5(\emptyset)$.
- For every $i \leq k$, $\phi_{i,1} \in \mathcal{L}^{n-1}(\Phi)$ (possibly $\psi' \wedge \neg\psi'$), and for $2 \leq j \leq l_i$, there exists $\phi'_{i,j} \in \mathcal{L}^{n-1}(\Phi)$ such that either $\phi_{i,j} = K\phi'_{i,j}$, or $\phi_{i,j} = \neg K\phi'_{i,j}$.

To prove that A_3–A_5 for ϕ are in $T_5(\mathcal{L}^{n-1}(\Phi))$, it is enough to prove that A_3–A_5 for $(\phi_{1,1} \vee \cdots \vee \phi_{1,l_1}) \wedge \cdots \wedge (\phi_{k,1} \vee \cdots \vee \phi_{k,l_k})$ are in $T_5(\mathcal{L}^{n-1}(\Phi))$. By successive applications of Lemma 3, it is enough to prove that, for every $i \leq k$, $(\phi_{i,1} \vee \cdots \vee \phi_{i,l_i})$ is in $T_5(\mathcal{L}^{n-1}(\Phi))$. We prove by induction on j that for every $j \leq l_i$, A_3–A_5 for $(\phi_{i,1} \vee \cdots \vee \phi_{i,j})$ are in $T_5(\mathcal{L}^{n-1}(\Phi))$. This is true for $j = 1$ since $\phi_{i,1} \in \mathcal{L}^{n-1}(\Phi)$. Assume this is true for j, then, by Lemma 5, A_3–A_5 for $(\phi_{i,1} \vee \cdots \vee \phi_{i,j}) \vee \phi_{i,j+1}$ are in $T_5(\{\phi_{i,1} \vee \cdots \vee \phi_{i,j}, \phi_{i,j+1}, \psi' \wedge \neg\psi'\}) \subseteq T_5(\mathcal{L}^{n-1}(\Phi))$. □

Proof of the Main Theorem. It follows from Lemma 6 that for every n, $T_5(\mathcal{L}^n(\Phi)) = T_5(\mathcal{L}^{n-1}(\Phi)) = T_5^0$. Since $T_5 = \cup_n T_n(\mathcal{L}^n(\Phi)) = T_5^0$, $T_5 = T_5^0$. □

3 Discussion

3.1 Partitional information structures

The standard modal logic system $S5$ is syntactic, in that it considers propositions and a knowledge operator. In order to make the connection between $S5$ (or $S5^0$) and partitional information, one needs to introduce the semantic representation, given by states of the world.

The bridge between semantic and syntactic models consists of Kripke structures (Kripke 1959), given as tuples $M = (\Omega, \pi, \mathcal{K})$; Ω is the set of states of nature; π :

$\Omega \times \Phi \to \{0, 1\}$ is a function assigning a truth value to every primitive proposition, i.e., $\pi(\omega, p) = 1$ if and only if p is true at ω; $\mathcal{K} : \Omega \to 2^{\Omega} \setminus \{\emptyset\}$ determines a binary relationship on Ω, often called the agent's possibility correspondence, i.e., $\omega' \in \mathcal{K}(\omega)$ means that the agent deems the state ω' possible while being at ω. We write $(M, \omega) \vDash \phi$ whenever ϕ is true at ω in the Kripke structure M. Truth is defined inductively in M at every state as follows:

$(M, \omega) \vDash p$ for each $p \in \Phi$ if and only if $\pi(\omega, p) = 1$
$(M, \omega) \vDash \phi$ if and only if $(M, \omega) \nvDash \neg\phi$
$(M, \omega) \vDash \phi \wedge \psi$ if and only if $(M, \omega) \vDash \phi$ and $(M, \omega) \vDash \psi$
$(M, \omega) \vDash K\phi$ if and only if $(M, \omega') \vDash \phi$ for all $\omega' \in \mathcal{K}(\omega)$

Recall that M is reflexive whenever $\omega \in \mathcal{K}(\omega)$ for all $\omega \in \Omega$; it is transitive whenever for all $\omega, \omega' \in \Omega$, if $\omega' \in \mathcal{K}(\omega)$ then $\mathcal{K}(\omega') \subseteq \mathcal{K}(\omega)$; finally, it is Euclidean whenever for all $\omega, \omega' \in \Omega$, if $\omega' \in \mathcal{K}(\omega)$ then $\mathcal{K}(\omega') \supseteq \mathcal{K}(\omega)$. Partitional information structures correspond to Kripke structures that are reflexive, transitive and Euclidean.

A proposition ϕ is a tautology in M, and we write $M \vDash \phi$, whenever $(M, \omega) \vDash \phi$ for all $\omega \in \Omega$. It is valid in a class of Kripke structures $\mathcal{M}$, and we write $\mathcal{M} \vDash \phi$, whenever ϕ is a tautology in every $M \in \mathcal{M}$.

An ML system is a sound axiomatization of a class of Kripke structures $\mathcal{M}$ whenever every theorem in this ML system is a valid proposition in $\mathcal{M}$. A modal logic system is a complete axiomatization of $\mathcal{M}$ whenever every valid proposition in $\mathcal{M}$ is a theorem in the ML system. It is well known that $S5$ is a sound and complete axiomatization of partitional Kripke structures (see e.g.,Chellas 1980; Lismont and Mongin 1994).

It immediately follows from our Main Theorem that $S5^0$ is also a sound and complete axiomatization of partitional Kripke structures.

3.2 Tightness of the result

We show that truth and introspection for the non-epistemic propositions are indispensable axioms, i.e., A_3–A_5 cannot be proven only from A_1–A_2 and R_1–R_2, using semantic models as introduced in the previous subsection.

Suppose, for instance, that there is a unique atomic proposition, $\Phi = \{p\}$, and consider a Kripke structure M such that $\Omega = \{\omega, \omega'\}$, with $\pi(\omega, p) = 1$ and $\pi(\omega', p) = 0$. Moreover, let $\mathcal{K}(\omega) = \{\omega\}$ and $\mathcal{K}(\omega') = \{\omega, \omega'\}$. It is known that M belongs to the class of Kripke structures which are axiomatized by A_1–A_4 together with the inference rules[2] R_1–R_2. Therefore, $\neg Kp \to K\neg Kp$ cannot be proven in M, implying that without assuming introspection for the non-epistemic propositions, we may obtain a strictly coarser set of theorems compared to $S5$.

Likewise, unless the truth axiom is assumed for the non-epistemic propositions, it cannot be proven by the remaining axioms. Consider, for instance, the following Kripke structure, M', such that $\Omega = \{\omega, \omega'\}$, with $\pi(\omega, p) = 1$ and $\pi(\omega', p) = 0$, and $\mathcal{K}(\omega) = \{\omega'\}$ and $\mathcal{K}(\omega') = \{\omega\}$. It is known that M' belongs to the class of Kripke structures which are axiomatized by A_1–A_2, A_4–A_5 together with the inference rules R_1–R_2 (Fagin et al. 1995), implying that $Kp \to p$ cannot be proven in M'.

[2] This implicit in Chellas (1980) and explicit in Lismont and Mongin (1994).

3.3 Epistemic depth

A well-known result in the literature states that in $S5$ every proposition is equivalent to a proposition of epistemic depth at most 1 (Hughes and Cresswell 1968, p. 51). In other words, for every $\phi \in \mathcal{L}(\Phi)$ there is some $\psi \in \mathcal{L}^1(\Phi)$ such that $\phi \leftrightarrow \psi$ is a theorem in $S5$. In this section we explain why the proof of our main theorem does not follow this result.

First, observe that $S5$ is a ML system with a larger (with respect to inclusion) set of axioms than $S5^0$, and therefore T_5^0 is a weak subset of T_5. Hence, there is in principle no reason why every proposition in Φ should be equivalent to a proposition in $\mathcal{L}^1(\Phi)$ in $S5^0$.[3]

Second, the known proofs the aforementioned standard result, such as in Hughes and Cresswell (1968) or Aumann (1999), make use of Theorems in $S5$ which are *a priori* not Theorems in $S5^0$. Of course, this is legitimate while working in $S5$, but isn't in $S5^0$. For instance, Aumann's proof (Lemma A.44) relies on Lemma A.41, which states that $K(K\phi \vee \psi) \leftrightarrow (K\phi \vee K\psi)$ is a theorem in $S5$. The proof of this lemma makes use of the fact that the truth axiom holds for all propositions and therefore applies A_3 for $(K\phi \vee \psi)$. Obviously, this cannot be done in $S5^0$, without first having proved that $K(K\phi \vee \psi) \rightarrow (K\phi \vee \psi)$ is a theorem in this system. Likewise, our proof significantly differs from the one by Hughes and Cresswell (1968), in that they in their definition of $S5$ they include a third inference rule, that of uniform replacing, according to which in every theorem of $S5$, if we replace a primitive proposition with any other proposition, the obtained proposition will also be a theorem. Though this inference rule clearly can be used in $S5$, it is not straightforward without using our main result that, if added in $S5^0$, it would not modify the set of theorems, T_5^0.

3.4 Interactive knowledge

Let us extend our analysis to a framework with multiple agents $\{1, \dots, n\}$, with typical elements i and j. Consider a separate knowledge modality K_i for every $i \in \{1, \dots, n\}$, implying that the set of all propositions, denoted by $\mathcal{L}_n(\Phi)$, now becomes the closure of Φ with respect to $\neg$, $\wedge$ and $K_1, \dots, K_n$, i.e., the language is enriched in order to contain propositions of the form "j knows that i knows p".

The ML system, $S5_n$, extends $S5$ to a multi-agent environment in which all axioms of $S5$ hold for each agent separately. For instance, each agent's knowledge satisfies the truth axiom, i.e., $K_i\phi \rightarrow \phi$ is an axiom for all ϕ and every i.

Recall the definition of the non-epistemic propositions in the single-agent framework: It is the set of sentences that do not contain any knowledge operator. Extending this definition to the multi-agent framework should be done with caution, e.g., consider the proposition "j knows that i knows p": From j's point of view, $K_j K_i p$ is epistemic, as it describes j's knowledge; on the other hand, $K_j K_i p$ is a non-epistemic proposition in i's language as it describes j's mental state, even though the latter refers

[3] It is, of course, a consequence of our Main Theorem that is actually the case.

to i's knowledge. Hence, from i's point of view any proposition starting with K_j is non-epistemic.

Formally, we define the set of non-epistemic propositions in i's language

$$\mathcal{L}_i^0(\Phi) := \mathcal{L}^0\left(\bigcup_{j\neq i}\left\{ K_j\phi \mid \phi \in \mathcal{L}(\Phi) \right\} \cup \Phi\right)$$

as the closure of $\bigcup_{j\neq i}\{K_j\phi \mid \phi \in \mathcal{L}(\Phi)\} \cup \Phi$ with respect to $\neg$ and $\wedge$. Obviously, when i is the only agent, it follows that $\mathcal{L}_i^0(\Phi) = \mathcal{L}^0(\Phi)$, implying that our generalized definition of non-epistemic propositions in the multi-agent environment is consistent with the single-agent case presented in the previous section.

Let $S5_n^0$ be the multi-agent generalization of $S5^0$:

A_1. All tautologies of propositional calculus
A_2. $\big(K_i\phi \wedge K_i(\phi \to \psi)\big) \to K_i\psi$
A_3. $K_i\phi \to \phi$, for all $\phi \in \mathcal{L}_i^0(\Phi)$
A_4. $K_i\phi \to K_iK_i\phi$, for all $\phi \in \mathcal{L}_i^0(\Phi)$
A_5. $\neg K_i\phi \to K_i\neg K_i\phi$, for all $\phi \in \mathcal{L}_i^0(\Phi)$
R_1. From ϕ and $(\phi \to \psi)$ infer ψ
R_2. From ϕ infer $K_i\phi$

Let T_{5n}^0 denote the theorems in $S5_n^0$.

Proposition 1 $T_{5n} = T_{5n}^0$.

Proof Observe that $\mathcal{L}(\Phi)$ coincides with the closure of $\mathcal{L}_i^0(\Phi)$ with respect to $\neg$, $\wedge$ and K_i. Moreover, similarly to the Main Theorem, we show that i can prove A_3–A_5 for all propositions in the closure of $\mathcal{L}_i^0(\Phi)$ with respect to $\neg$, $\wedge$ and K_i, and therefore i can prove A_3–A_5 for all propositions in $\mathcal{L}(\Phi)$. Likewise, for every individual which completes the proof. □

It follows directly, from the previous result, that $S5_n^0$ is a sound and complete axiomatization of the class of multi-agent partitional Kripke structures.

Notice that in order to prove A_3–A_5 for all propositions it does not suffice to assume the truth axiom and introspection only for $\mathcal{L}^0(\Phi)$, e.g., even if j assumes $(K_i\phi \to \phi)$, he cannot prove $(K_jK_i\phi \to K_i\phi)$. The reason is that, from j's point of view, $K_i\phi$ is a non-epistemic proposition, and therefore j cannot infer the truth axiom for $K_i\phi$.

Acknowledgments We are indebted to Geir Asheim, Giacomo Bonanno, Amanda Friedenberg, Willemien Kets, Friederike Mengel, Andrés Perea, Philippe Mongin, Mark Voorneveld, and Jörgen Weibull for fruitful discussions and useful comments. We would also like to thank the audiences of the Epistemic Game Theory Workshop (Stony Brook), Game Theory Conference (Stony Book), ESEM (Barcelona), SAET (Ischia), CRETE (Tinos), Bocconi University, LSE, Santa Fe Institute, Paris Game Theory Seminar, IHPST (Paris), Stockholm School of Economics, University of Warwick, HEC Lausanne, and Göteborg University. The financial support from the Marie Curie Intra-European Fellowship is gratefully acknowledged (PIEF-GA-2009-237614).

References

Aumann, R. J. (1976). Agreeing to disagree. *Annals of Statistics, 4*, 1236–1239.

Aumann, R. J. (1987). Correlated equilibrium as an expression of Bayesian rationality. *Econometrica, 55*, 1–18.

Aumann, R. J. (1999). Interactive epistemology I: Knowledge. *International Journal of Game Theory, 28*, 263–300.

Aumann, R. J., & Brandenburger, A. (1995). Epistemic conditions for Nash equilibrium. *Econometrica, 63*, 1161–1180.

Chellas, B. F. (1980). *Modal logic: An introduction*. Cambridge: Cambridge University Press.

Fagin, R., Halpern, J. Y., Moses, Y., & Vardi, M. Y. (1995). *Reasoning about knowledge*. Cambridge: MIT press.

Geanakoplos, J. (1989). Game theory without partitions, and applications to speculation and consensus. *Cowles Foundation Discussion Paper* No. 914.

Harsanyi, J. (1967–1968). Games with incomplete information played by Bayesian players, I–III. *Management Science*, *14*, 159–182, 320–334, 486–502.

Hughes, G. E., & Cresswell, M. J. (1968). *An introduction to modal logic*. London: Methuen.

Kripke, (1959). A completeness theorem in modal logic. *Journal of Symbolic Logic, 24*, 1–14.

Lismont, L., & Mongin, P. (1994). On the logic of common belief and common knowledge. *Theory and Decision, 37*, 75–106.

Samet, D. (1990). Ignoring ignorance and agreeing to disagree. *Journal of Economic Theory, 52*, 190–207.