UNIVERSITY OF CAPE TOWN

An epistemic model of an agent who does not reflect on reasoning processes

Alexander Zimper[1]

Working Paper Number 45

[1] School of Economics, University of Cape Town

An epistemic model of an agent who does not reflect on reasoning processes*

Alexander Zimper[†]

Revised version: December 13, 2006

Abstract

This paper introduces an epistemic model of a boundedly rational agent under the two assumptions that (i) the agent's reasoning process is in accordance with the model but (ii) the agent does not reflect on these reasoning processes. For such a concept of bounded rationality a semantic interpretation by the *possible world semantics* of the Kripke (1963) type is no longer available because the definition of knowledge in these possible world semantics implies that the agent knows all valid statements of the model. Key to my alternative semantic approach is the extension of the method of truth tables, first introduced for the propositional logic by Wittgenstein (1922), to an epistemic logic so that I can determine the truth value of epistemic statements for all relevant truth conditions. I also define an axiom system plus inference rules for knowledge- and unawareness statements whereby I drop the inference rule of *necessitation*, which claims that an agent knows all theorems of the logic. As my main formal result I derive a determination theorem linking my semantic with my syntactic approach.

Keywords: Bounded Rationality, Knowledge, Unawareness, Epistemic Logic, Semantic Interpretation, Iterative Solution Concepts for Strategic Games

JEL Classification Numbers: B41, C72, D83.

*I thank Itzhak Gilboa, Juliane Hiesgen, Alexander Ludwig and two anonymous referees for helpful comments and suggestions.

[†]School of Economics, University of Cape Town, Private Bag, Rondebosch, 7701, Cape Town, South Africa. Email: zimper@bigfoot.com

1 Introduction

Existing epistemic models consider agents who know all theorems of the underlying logical language. Such an agent would, for example, know all tautologies of the propositional calculus; however, this is impossible for any real human being. While these existing epistemic models thus describe the benchmark case of a super-rational individual who is perfectly reflective about the - supposed - rules of human reasoning processes, this paper introduces the opposite benchmark case of an individual who does not reflect at all on the rules of human reasoning processes. Such an individual is still a *logical* agent in the sense that his reasoning processes are comprehensively described by the theorems of the considered epistemic logic. He is, however, also *boundedly rational* in the sense that he does not know these theorems. The epistemic model developed in this paper is - arguably - closer to reality than existing epistemic models since hardly any real human reflects on how they or others arrive at logical conclusions.

If an agent does not reflect on reasoning processes, a semantic interpretation by *possible world semantics* of the Kripke (1963) type is no longer available because the definition of knowledge in these possible world semantics implies that the agent knows all valid statements of the model. As this paper's **main conceptual contribution** I therefore introduce a semantic interpretation of epistemic statements under the assumption that an agent does not necessarily know the valid statements of the model.[1] Key to my alternative semantic approach, called ESM, is the development of *truth tables* which determine the truth value of epistemic statements for all relevant truth conditions. Furthermore, I define an epistemic logic, i.e., an axiom system plus inference rules for knowledge- and unawareness statements, called TKU, where I drop the inference rule of *necessitation*, which claims that an agent knows all theorems of the logic. As my main formal result I derive a determination theorem which states that every theorem of TKU is a tautology of ESM and that every tautology of ESM is a theorem of TKU.

As a **secondary conceptual contribution**, the present paper provides a semantic interpretation of unawareness statements by means of truth tables. In the economic literature (cf. Theorem 1.i in Dekel, Lipman, and Rustichini, 1998a) it is well known that the inference rule of necessitation is incompatible with a non-degenerate concept of unawareness. Subsequent research (cf. Section 2.3: "Related literature") has therefore focussed on the development of epistemic models which may incorporate unawareness statements. Since I do not impose necessitation as a valid inference rule, my epistemic

[1] Notice that the agent's *inability to reflect on reasoning processes* refers to the situation where the agent does not know the valid statements of the model. In contrast, the *inability for "positive introspection"*, discussed in the epistemic literature, describes the situation where the agent does not know that he knows some statement.

model of a boundedly rational agent can easily accommodate such a non-degenerate concept of unawareness.

The remainder of the paper is organized as follows. Section 2 explains in more detail my semantic as well as my syntactic approach; moreover, the relation of my approach to the existing literature is discussed.. Section 3 introduces the formal language. In Section 4 the iterative procedure for determining the truth value of formulas is presented; with examples given in Section 5. I describe my syntactic approach in Section 6 where I also present the determination theorem. The according proofs of soundness and completeness are relegated to the Appendix. As an application of my epistemic logic I argue in Section 7 that the standard argument of "common knowledge of rationality" may not be sufficient for the epistemic justification of iterative solution concepts for strategic games. Section 8 concludes with a critical discussion of my approach.

2 The approach

2.1 The semantic approach

A semantic model of the Kripke (1963) type, i.e., a *standard model*, is defined as a triple $\mathcal{M} = \langle W, R, P \rangle$ such that W is a set of possible worlds, R is a binary relation on W (interpreted as the "possibility"/"reachability" relation between different worlds), and P is a mapping that assigns to every atomic sentence of the logic the set of possible worlds at which this atomic sentence is true. According to the semantic interpretation of the knowledge operator by a standard model, it is true that the agent knows statement A in world $\alpha \in W$ if and only if statement A is true in all possible worlds β that are reachable from α, i.e., $\alpha R \beta$. Since tautologies are, by definition, true in **all** possible worlds, any tautology of a standard model is known to the agent. As a consequence, standard models are not suitable for a semantic interpretation of epistemic statements about an agent who is boundedly rational in my sense. I therefore use the traditional method of truth tables for the semantic interpretation of epistemic statements.

The semantic interpretation by truth tables (introduced by Wittgenstein, 1922, paragraphs 4.25-4.53) is standard for propositional statements. Consider some atomic sentence p (e.g., *T. Pynchon is a reclusive*) and observe that there are exactly two truth conditions (=*states of affair*) - namely p *is true* (=v) and p *is false* (=f) - that are relevant to the semantic interpretation of p. The corresponding truth table looks as follows:

p
v
f

For the semantic interpretation of more complex propositional formulas, which include k different atomic sentences, there are exactly 2^k possible combinations of the atomic sentences' truth values. For example, there are exactly four truth conditions that are relevant to the semantic interpretation of a statement $p \Rightarrow q$ (e.g., **IF** *T. Pynchon is a reclusive* **THEN** *J.D. Salinger is a reclusive*) so that $p \Rightarrow q$ is false iff p is true whereas q is false. As a truth table one therefore obtains:

p	$\Rightarrow$	q
v	v	v
v	f	f
f	v	v
f	v	f

For epistemic statements, i.e., statements that include knowledge or unawareness operators, the development of truth tables is not as straightforward since there are typically more relevant truth conditions than all possible combinations of truth values for atomic sentences appearing in the statement. For example, in my interpretation of the knowledge operator we want to ensure that an epistemic statement of the form *Agent X knows that T. Pynchon is a reclusive*, denoted as $K(p)$, must be *false* when it is *false* that *T. Pynchon is a reclusive* whereas it may be either *true* or *false* when it is *true* that *T. Pynchon is a reclusive.* As a consequence, there are three different truth conditions which are relevant to the semantic interpretation of statement $K(p)$ so that the corresponding truth table writes as:

K	(p)
v	v
f	v
f	f

Similarly, for the semantic interpretation of the epistemic statement *Agent X is unaware that T. Pynchon is a reclusive*, denoted as $U(p)$, I assume that regardless whether it is *true* or *false* that *T. Pynchon is a reclusive* it might be *true* or *false* that the agent is totally unaware of the possibility that there might be some reclusive subject identified as T. Pynchon. Hence, there are four relevant truth conditions for statement $U(p)$ so that the truth table is given as:

U	(p)
v	v
f	v
v	f
f	f

More specifically, besides adopting the standard semantics for propositional operators, I stipulate for ESM the following truth conditions for the knowledge operator, K, and the unawareness operator, U:

Knowledge operator

1. If A is true, then the agent may or may not know A. That is, if A is true, then $K(A)$ may be true or false.

2. If A is false, then the agent must not know A, i.e., $K(A)$ must be false if A is false.

3. If the agent knows that A implies B, then the agent knows B if he knows A. Formally, if $K(A \Rightarrow B)$ and $K(A)$ are true, then $K(B)$ is also true.

Unawareness operator

4. If it is true that the agent is unaware of A, i.e., if $U(A)$ is true, then A may be true or false.

5. If it is false that the agent is unaware of A, i.e., if $U(A)$ is false, then A may be true or false.

6. If the agent is unaware of A, then it is false

- that the agent knows A;
- that he knows that he does not know A; etc.

That is, if $U(A)$ is true, then $K(A)$, ..., $K(\neg K(....\neg K(A)))$, etc. must be false.

7. If the agent is unaware of A, then it is false

- that the agent knows his unawareness;
- that he knows that he does not know his unawareness; etc.

That is, if $U(A)$ is true, then $K(U(A))$, ..., $K(\neg K(....\neg K(U(A))))$, etc. must be false.

In Section 4 I describe an iterative procedure that unambiguously determines the truth table for any given formula of the language (introduced in Section 3) according to these semantic conditions. Through this procedure it is therefore possible to decide for any formula whether it is a tautology (*true* for all relevant truth conditions), a contradiction (*false* for all relevant truth conditions), or an empirical statement (neither *true* nor *false* for all relevant truth conditions) of ESM.

2.2 The syntactic approach

While the semantic approach looks at conditions that determine the truth value of statements, the syntactic approach defines an *epistemic logic* as some finite collection of axioms and inference rules. The theorems of this logic (i.e., any statement that may be derived from these axioms by the application of inference rules) may be interpreted as the rules by which an agent processes knowledge and arrives at conclusions. Epistemic logics are typically defined as some *modal logic*, e.g., the axiom systems S5, KD45 etc. (cf. Chellas, 1980), for which the *necessity operator* $\Box$ is interpreted as the knowledge operator K. While, in general, some statement A may be either true or false, the modal statement $\Box(A)$ qualifies A as a statement that is necessarily true. Given that the theorems of some modal logic are interpreted as logical laws, it is natural to claim that all theorems of this logic must necessarily be true. For modal logics this claim is formally accomplished by the inference rule of necessitation, which states that we can deduce $\Box(A)$ as a theorem whenever A is a theorem. However, when an epistemic logic is just defined as such a modal logic, the inference rule of necessitation implies that the agent has absolute power of reflection on reasoning processes in the sense that he knows every theorem of the epistemic logic.

Applied to an epistemic logic, I therefore regard necessitation (also called *knowledge generalization*) as an extremely idealizing assumption. In my opinion, any relevant concept of a boundedly rational agent has to drop - or at least to weaken - necessitation as a valid inference rule. Since I am interested in a boundedly rational agent who might not have any power of self-reflection, I introduce an epistemic logic, called TKU, which excludes necessitation as inference rule without substituting it by some weakened version. More specifically, TKU will be defined as follows (for the moment being just read the symbols $\Rightarrow, \neg$ as "IF...THEN", "NOT"):

All theorems and inference rules of propositional logic are theorems and inference rules of TKU;

The *knowledge axiom* K, i.e., $K(p) \Rightarrow p$, and the *knowledge distribution axiom* T, i.e., $K(p \Rightarrow q) \Rightarrow (K(p) \Rightarrow K(q))$, are axioms of TKU;

With respect to the relationship between the unawareness- and the knowledge operator, I assume that $\neg K(U(p))$, $U(p) \Rightarrow \neg K(p)$ and $U(p) \Rightarrow \neg K(U(p))$ are axioms of TKU. For example, the following statements are theorems of TKU: *Agent X* *does not know that he is unaware that T. Pynchon is a reclusive*; and *If Agent X is unaware that T. Pynchon is a reclusive then he does neither know that T. Pynchon is a reclusive nor does he know that he is unaware that T. Pynchon is a reclusive.*

Regarding the relationship between the unawareness- and the knowledge operator, I also introduce two epistemic inference rules of TKU, which can be interpreted as formalizations of the assumption that unawareness implies arbitrary levels of ignorance such as *Agent X does not know that he does not know that T. Pynchon is a reclusive, if he is unaware that T. Pynchon is a reclusive*; and *Agent X does not know that he does not know that he is unaware that T. Pynchon is a reclusive, if he is unaware that T. Pynchon is a reclusive.*

2.3 Related literature

As the present paper, Fagin and Halpern (1988) depart from the observation that a semantic interpretation by standard models only applies to epistemic statements of agents which satisfy extremely strong rationality assumptions. For the semantic interpretation of epistemic statements of a boundedly rational agent, Fagin and Halpern propose so-called *awareness structures* $\mathcal{M}^* = \langle W, R, P, \mathcal{A} \rangle$ which extend standard models by the *awareness-operator* $\mathcal{A}$ that assigns to every possible world $\alpha \in W$ some subset of the formulas in P. According to their interpretation, the set $\mathcal{A}(a) \subset P$ contains all the statements the agent is aware of in world α. According to the semantic interpretation of the knowledge operator by an awareness structure, it is true that the agent knows statement A in world $\alpha \in W$, if and only if, (i) statement A is true in all possible worlds β that are reachable from α, i.e., $\alpha R \beta$, and (ii) the agent is aware of statement A in world $\alpha \in W$, i.e., $A \in \mathcal{A}(\alpha)$. Formally, the appropriate specification of $\mathcal{A}$ may therefore determine any subset of statements known to the agent according to the standard model $\mathcal{M} = \langle W, R, P \rangle$ as the set of statements known to the agent according to the awareness structure $\mathcal{M}^* = \langle W, R, P, \mathcal{A} \rangle$.

In order to avoid replacing the strong semantic interpretation of the knowledge operator by standard models by an arbitrary principle, Fagin and Halpern discuss plausible properties of the awareness operator $\mathcal{A}$ which effectively restrict the arbitrariness of knowledge in awareness-structures. Among other proposals, Fagin and Halpern consider the case where awareness is generated by atomic sentences, that is, the agent is aware of a statement, if and only if, he is aware of all atomic sentences that appear in this statement. Moreover, Halpern (2001) observes that awareness-structures with this particular specification of the awareness operator are basically equivalent to so-called *generalized standard models* of Modica and Rustichini (1999). Modica and Rustichini derive a corresponding determination theorem for an epistemic logic where the inference rule of necessitation is reduced to an inference rule implying that whenever some atomic sentences are known to the agent then he must also know all theorems that (exclusively) refer to these atomic sentences.

While the proposals of Fagin and Halpern and of Modica and Rustichini for controlled generalizations of standard models are very compelling, they still impose rather strong rationality assumptions. They describe an agent with absolute power of self-reflection restricted to the domain of statements he is aware of, so that the agent knows, e.g., all tautologies containing atomic sentences he is aware of (Modica and Rustichini, 1999). In contrast, this paper describes a boundedly rational agent who might not have any power of self-reflection. Therefore, my syntactic approach TKU simply drops necessity as a valid inference rule without substituting it, as in Modica and Rustichini, by some weakened version restricted to some domain of statements the agent is aware of.

Recently, economic theorists have become interested in epistemic logics where agents may be *unaware* about statements (=events) in the sense that (cf. Modica and Rustichini, 1994; Dekel, Lipman, and Rustichini, 1998a, 1998b):

An agent does not know some statement,

and he does not know that he does not know this statement,

and so on...

Building on the seminal contribution of Modica and Rustichini (1994), Dekel, Lipman, and Rustichini (1998a) claim that the following three axioms stand for natural properties of an according unawareness operator:

"KU-introspection": $\neg K(U(A))$

"Plausibility": $U(A) \Rightarrow (\neg K(A) \wedge \neg K(\neg K(A)))$

"AU-introspection": $U(A) \Rightarrow U(U(A))$

where A is an arbitrary statement of the language. As their main result, Dekel, Lipman, and Rustichini demonstrate (Theorem 1.i) that for any epistemic logic which satisfies these three axioms "the agent is never unaware of anything" (Dekel, Lipman, and Rustichini, 1998a p. 166) when necessitation is a valid inference rule of this logic. The proof is easy: Applying necessitation to KU-intropspection gives $K(\neg K(U(A)))$ whereas plausibility and AU-introspection imply $U(A) \Rightarrow \neg K(\neg K(U(A)))$, which results in a contradiction whenever $U(A)$ is a true statement, i.e., whenever the agent is unaware of anything. As a consequence, Dekel, Lipman, and Rustichini conclude that the inference rule of necessitation precludes any non-trivial epistemic logic with unawareness.

Similarly, my epistemic logic TKU contains KU-introspection, $\neg K(U(p))$, as an axiom and the statement $U(p) \Rightarrow \neg K(\neg K(U(p)))$ follows as theorem of TKU. Thus, if necessitation was a valid inference rule of TKU, I would run into the same difficulty with unawareness statements as observed by Dekel, Lipman, and Rustichini. However, since TKU precludes necessitation as a valid inference rule, my boundedly rational agent may be actually unaware about something without running into contradictions.

An alternative approach to modelling unawareness is proposed in Heifetz, Meier, and Schipper (2005) who proceed in the tradition of Aumann's (1976) set-theoretic notion of the knowledge operator in order to develop a generalized structure of a state space which admits for non-trivial unawareness statements. Key to their approach is the definition of the whole state space as a union of different state spaces that are partially ordered with respect to their "expressiveness". The proposed semantic interpretation of states of the world as truth-value combinations in example 1 (Heifetz at al., 2005) seems to suggest that there might be a close link between their approach and a semantic interpretation by truth tables with respect to three different truth values for atomic sentences. For instance, the different states of the world that define in their example 1 the event *Statement* r *is true* could be equivalently written as all truth value combinations where r is assigned v given that the truth values of the atomic sentences p, q, r can take on values in $\{v, f, ?\}$ where ? indicates that the statement is neither *true* nor *false*. And indeed, in a recent investigation Halpern and Rego (2005) demonstrate that the approach of Heifetz at al.(2005) "[...] differs from standard epistemic logic by allowing a third truth value."(p. 11).

While the present paper considers an agent who is boundedly rational in the sense that he does not necessarily reflect on the logical rules he applies, Kaneko and Suzuki (2003, 2005) are interested in a concept of bounded rationality which captures the limited ability of agents to apply logical rules. In particular, they propose a measure for the complexity of logical deductions within an inference system of intuitionistic logic of the Gentzen-type so that the proof of a particular formula may demand more rationality on behalf of the agent than the proof of another formula. In contrast, my approach presumes that the imposed logical rules correctly describe the reasoning process of the agent no matter how complex these rules are.

3 Preliminaries: The formal language

This section defines our formal language as a collection of *well formulated formulas (=wff)*, which may differ by their *epistemic degree*. Formally, I partition the set of all wff into the subsets Λ^k, $k \in \{0, 1, 2, ...\}$, where k stands for the epistemic degree of the wff belonging to Λ^k.

Definition: *Wff (well formulated formulas) are iteratively constructed, for* $k \in \{0, 1, 2, ...\}$*, as follows:*

0.1 *Variables for atomic sentences, e.g.,* $p, q, r, p^1, p^2, ...$ *are members of* Λ^0*.*

0.2 *If* $A, B \in \Lambda^0$ *then* $\neg A, (A \Rightarrow B) \in \Lambda^0$*.*

0.3 *All wff* $A \in \Lambda^0$ *are derived from 0.1 and 0.2.*

k+1.1 *If* $A \in \Lambda^k$ *then* $K(A), U(A) \in \Lambda^{k+1}$*.*

k+1.2 *If* $A \in \Lambda^{k+1}$ *and* $B \in \Lambda^j$ *with* $j \leq k+1$*, then* $\neg A, (A \Rightarrow B), (B \Rightarrow A) \in \Lambda^{k+1}$.

k+1.3 *All wff* $A \in \Lambda^{k+1}$ *are derived from k+1.1 and k+1.2.*

The symbol $\neg$ denotes the *negation* - to be read as: NOT - of the classical propositional logic; $\Rightarrow$ denotes the *subjunction* - to be read as: IF ... THEN. K is interpreted as *knowledge operator* so that $K(A)$ reads as "*The agent knows* A." U is interpreted as *unawareness operator* so that $U(A)$ reads as "*The agent is unaware of* A." The above definition of a wff entails that any wff can be constructed in an obvious way by an iterative procedure that starts with atomic sentences and either links existing wff by the connective $\Rightarrow$ or adds $\neg, K, U$ to existing wff. Whenever some wff A appears at some stage in the iterative procedure which generates a wff C, we say that *wff* C *contains wff* A or equivalently: *wff* A *occurs in wff* C. Λ^0 collects all wff of the propositional calculus. Furthermore, observe that a wff C belongs to Λ^k with $k \geq 1$ if and only if there occur $K(A)$ or $U(B)$ in C such that A and B belong to Λ^{k-1}. Thus, the epistemic degree of a wff is a measure for the "nestedness" of epistemic operators appearing in the wff. The epistemic wff $K(A)$ or $U(B)$ occur *separated* in wff C whenever they are not contained in some epistemic wff that also occurs in C. For example, $U(p)$ and $K(U(q))$ are separated in $K(U(q)) \Rightarrow U(p)$ but $U(q)$ is not.

Note that all wff which are not atomic sentences can be either described as $\neg A$, $A \Rightarrow B$ (whereby we drop the outer parentheses), $K(A)$ or $U(A)$. The respective operators $\neg$, $\Rightarrow$, K or U will be called the *main operators* of these wff. Given some wff F such that

$$F := A^1 \Rightarrow \left(A^2 \Rightarrow \left(... \Rightarrow \left(A^k \Rightarrow E\right) ...\right)\right),$$

we call the wff $A^1, A^2, ..., A^k$ the *antecedents* and the wff E the *consequent* of F.

Remark: I consider only the propositional operators $\neg$ and $\Rightarrow$ because the formal proofs are much more transparent when the formal language is kept as lean as possible. Observe that alternative connectives of the propositional calculus such as $\wedge$, $\vee$, $\Leftrightarrow$ (to be read as: AND, OR, IF AND ONLY IF) could easily be introduced into our formal language by the following notational conventions:

$$\begin{aligned} \neg(A \Rightarrow \neg B) &= {}_{def} A \wedge B \\ (\neg A \Rightarrow B) &= {}_{def} A \vee B \\ \neg((A \Rightarrow B) \Rightarrow \neg(B \Rightarrow A)) &= {}_{def} (A \Leftrightarrow B) . \end{aligned}$$

4 An iterative procedure for the development of truth tables

This section presents an iterative procedure which effectively develops the truth table for any given wff such that the semantic conditions 1-7 of ESM are satisfied. Since the rows of such a truth table represent all truth conditions that are relevant for deciding whether the wff is *true* or *false*, we interpret these rows as the *relevant states of affair*.

Step 1. Write down in a table all possible combinations of truth value assignments, v or f, to atomic sentences that occur in E. Note that if there are k different atomic sentences, then this tableau will consist of 2^k rows of different truth value assignments.

Step 2. If some wff A, B, occurring in E, are assigned some truth value, then assign truth values to joint wff, occurring in E, according to the standard rules for propositional operators

A	$\neg A$
v	f
f	v

A	B	$A \Rightarrow B$
v	v	v
v	f	f
f	v	v
f	f	v

Step 3. If there occur epistemic wff $K(A)$ in E such that A has some truth value while $K(A)$ has not, then **duplicate** all rows of the table where v is assigned to A and assign truth values to $K(A)$ as follows:[2]

[2]It can easily be shown that the order of duplication does not matter if there are several wff $K(A_1), K(A_2), \ldots$ such that $A_1, A_2, \ldots$ have truth values whereas $K(A_1), K(A_2), \ldots$ have not. (The same observation applies to step 5.)

A	$K(A)$
v	v
v	f
f	f

Step 4. If there occur epistemic wff $K(A \Rightarrow B)$, $K(A)$, and $K(B)$ in E with truth values, then **delete** all rows from the table which contain the following truth values:

$K(A \Rightarrow B)$	$K(A)$	$K(B)$
v	v	f

Step 5. If there occur epistemic wff $U(A)$ in E such that A has some truth value while $U(A)$ has not, then **duplicate** all rows of the table and assign truth values to $U(A)$ as follows:

A	$U(A)$
v	v
v	f
f	v
f	f

Step 6. If there occur epistemic wff $U(A)$ and $K(U(A))$ or ... or $K(\neg K(....\neg K(U(A))))$ in E with truth values, then **delete** all rows such that $U(A)$ and $K(U(A))$ or ... or $U(A)$ and $K(\neg K(....\neg K(U(A))))$ are simultaneously true.

Step 7. If there occur epistemic wff $U(A)$ and $K(A)$ or ... or $K(\neg K(....\neg K(A)))$ in E with truth values, then **delete** all rows such that $U(A)$ and $K(A)$ or ... or $U(A)$ and $K(\neg K(....\neg K(A)))$ are simultaneously true.

Step 8. If E contains some wff without truth values repeat step 2 - step 7.

5 Examples: Tautologies and empirical statements

Given any wff E, the described iterative procedure assigns after a finite number of steps truth values to the main-operator of E so that there do not remain any wff occurring in E without a truth value. If these truth values for the main-operator of E are *true* (*false)* at all rows of the table we call E a *tautology* (*contradiction*) of ESM. Otherwise (i.e., there are *true* as well as *false* truth values) we speak of an empirical statement. That is, there exists some states of the world where an empirical statement is true while there also exist some states of the world where it is false. So, to decide whether such a statement is actually true or false is ultimately an empirical but not a logical task.

The following wff turn out to be **tautologies** of ESM:

1. $K(p) \Rightarrow p$
2. $\neg K(U(p))$
3. $K(p \Rightarrow q) \Rightarrow (K(p) \Rightarrow K(q))$
4. $K(p) \Rightarrow \neg K(\neg p)$
5. $K(p) \Rightarrow \neg U(p)$
6. $K(\neg p) \Rightarrow \neg U(p)$
7. $K(U(p)) \Rightarrow \neg U(p)$
8. $U(p) \Rightarrow \neg K(p)$

...

8+k. $U(p) \Rightarrow \neg K(\neg K(...\neg K(p)))$

9+k. $U(p) \Rightarrow \neg K(U(p))$

...

9+2k. $U(p) \Rightarrow \neg K(\neg K(...\neg K(U(p))))$;

whereas the following wff are only **empirical statements** of ESM:

1. $K(p) \Rightarrow K(K(p))$ (="positive introspection")
2. $\neg K(p) \Rightarrow K(\neg K(p))$ (= "negative introspection")
3. $U(p) \Rightarrow U(U(p))$
4. $U(U(p)) \Rightarrow U(p)$
5. $\neg K(\neg K(U(p)))$

...

5+k-1. $\neg K(\neg K(...\neg K(U(p))))$.

In what follows I prove this claim for selected wff by describing in detail the development of truth tables for these wff, (the proofs for $K(p) \Rightarrow p$ and $K(p \Rightarrow q) \Rightarrow (K(p) \Rightarrow K(q))$ appear in the Appendix where soundness is proved; the proofs for the remaining wff are left to the reader).

Example 1. $\neg K(U(p))$ is a tautology of ESM.

Proof:

1. Assign truth values to atomic sentences according to step 1:

$\neg$	K	$(U$	$(p))$
			v
			f

2. Duplicate all rows and assign truth values according to step 5:

$\neg$	K	$(U$	$(p))$
		v	v
		f	v
		v	f
		f	f

3. Duplicate rows where $U(p)$ is true and assign truth values according to step 3:

$\neg$	K	$(U$	$(p))$
	v	v	v
	f	v	v
	f	f	v
	v	v	f
	f	v	f
	f	f	f

4. According to step 8, delete the rows where $K(U(p))$ and $U(p)$ are simultaneously true:

$\neg$	K	$(U$	$(p))$
	f	v	v
	f	f	v
	f	v	f
	f	f	f

5. Finally, by step 2:

$\neg$	K	$(U$	$(p))$
v	f	v	v
v	f	f	v
v	f	v	f
v	f	f	f

In all rows the main-operator is assigned *true*, which proves the claim.□

Example 2. $U(p) \Rightarrow \neg K(p)$ is a tautology of ESM.

Proof:

1. Assign truth values to atomic sentences according to step 1:

U	(p)	$\Rightarrow$	$\neg$	K	(p)
	v				v
	f				f

2. Duplicate rows where $K(p)$ is true and assign truth values according to step 3:

U	(p)	$\Rightarrow$	$\neg$	K	(p)
	v			v	v
	v			f	v
	f			f	f

3. Duplicate all rows and assign truth values according to step 5:

U	(p)	$\Rightarrow$	$\neg$	K	(p)
v	v			v	v
f	v			v	v
v	v			f	v
f	v			f	v
v	f			f	f
f	f			f	f

4. According to step 7, delete all rows where $U(p)$ and $K(p)$ are simultaneously true:

U	(p)	$\Rightarrow$	$\neg$	K	(p)
f	v			v	v
v	v			f	v
f	v			f	v
v	f			f	f
f	f			f	f

5. By step 2:

U	(p)	$\Rightarrow$	$\neg$	K	(p)
f	v		f	v	v
v	v		v	f	v
f	v		v	f	v
v	f		v	f	f
f	f		v	f	f

5. Finally, by step 2:

U	(p)	$\Rightarrow$	$\neg$	K	(p)
f	v	**v**	f	v	v
v	v	**v**	v	f	v
f	v	**v**	v	f	v
v	f	**v**	v	f	f
f	f	**v**	v	f	f

In all rows the main-operator is assigned *true*, which proves the claim.□

Example 3. $K\,(p) \Rightarrow \neg K\,(\neg p)$ is a tautology of ESM.

Proof:

1. By step 1:

K	(p)	$\Rightarrow$	$\neg$	K	$(\neg$	$p)$
	v					v
	f					f

2. By step 2:

K	(p)	$\Rightarrow$	$\neg$	K	$(\neg$	$p)$
	v				f	v
	f				v	f

3. By step 3:

K	(p)	$\Rightarrow$	$\neg$	K	$(\neg$	$p)$
v	v				f	v
f	v				f	v
f	f				v	f

4. By step 3:

K	(p)	$\Rightarrow$	$\neg$	K	$(\neg$	$p)$
v	v			f	f	v
f	v			f	f	v
f	f			v	v	f
f	f			f	v	f

5. Repeating step 2 twice finally gives:

K	(p)	$\Rightarrow$	$\neg$	K	$(\neg$	$p)$
v	v	**v**	v	f	f	v
f	v	**v**	v	f	f	v
f	f	**v**	f	v	v	f
f	f	**v**	v	f	v	f

□

Example 4. $K(p) \Rightarrow K(K(p))$, i.e., "positive introspection", is not a tautology of ESM.

Proof:

1. By step 1:

K	(p)	$\Rightarrow$	K	$(K$	$(p))$
	v				v
	f				f

2. By step 3:

K	(p)	$\Rightarrow$	K	$(K$	$(p))$
v	v			v	v
f	v			f	v
f	f			f	f

3. By step 3:

K	(p)	$\Rightarrow$	K	$(K$	$(p))$
v	v		v	v	v
v	v		f	v	v
f	v		f	f	v
f	f		f	f	f

4. By step 2:

K	(p)	$\Rightarrow$	K	$(K$	$(p))$
v	v	**v**	v	v	v
v	v	**f**	f	v	v
f	v	**v**	f	f	v
f	f	**v**	f	f	f

□

6 The main result: A determination theorem

In this section I present the epistemic logic TKU, which I define as the axioms and inference rules of the propositional calculus **P2** (cf. Church, 1996, p.119ff.) **plus** two knowledge axioms, three unawareness axioms, and two additional inference rules for unawareness statements:

Axioms of P2

A1: $p \Rightarrow (q \Rightarrow p)$

A2: $(p \Rightarrow (q \Rightarrow r)) \Rightarrow ((p \Rightarrow q) \Rightarrow (p \Rightarrow r))$

A3: $(\neg p \Rightarrow \neg q) \Rightarrow (q \Rightarrow p)$

Knowledge axioms

T: $K(p) \Rightarrow p$

K: $K(p \Rightarrow q) \Rightarrow (K(p) \Rightarrow K(q))$

Unawareness axioms

U1: $\neg K(U(p))$

U2: $U(p) \Rightarrow \neg K(p)$

U3: $U(p) \Rightarrow \neg K(U(p))$

Inference rules of P2

Modus ponens:

$$\frac{A, (A \Rightarrow B)}{B}$$

Simultaneous substitution:

$$\frac{A}{A\{p_1, ..., p_m/C_1, ..., C_m\}}$$

where $A\{p_1,...,p_m/C_1,...,C_m\}$ denotes the wff which results when the atomic sentences p_j, $j=1,...,m$, occurring in A, are replaced by wff C_j , $j=1,...,m$.

Inference rules for Unawareness-Statements

UR1:

$$\frac{U(A)\Rightarrow\neg K(...\neg K(A))}{U(A)\Rightarrow\neg K(\neg K(...\neg K(A)))}$$

UR2:

$$\frac{U(A)\Rightarrow\neg K(...\neg K(U(A)))}{U(A)\Rightarrow\neg K(\neg K(...\neg K(U(A))))}$$

A wff A is called a *theorem* of the epistemic logic TKU if and only if it is either one of the axioms of TKU or it can be derived from these axioms by an application of the inference rules of TKU. Now we are ready to present this paper's main result, which states that the epistemic logic TKU is determined by the semantic model ESM.

Determination Theorem

- TKU *is sound with respect to* ESM*, i.e., every theorem of* TKU *is a tautology of* ESM*.*
- TKU *is complete with respect to* ESM*, i.e., every tautology of* ESM *is a theorem of* TKU*.*

The formal details of the proof are relegated to the appendix. Proving soundness is rather straightforward and standard: At first I show that all axioms of TKU are tautologies of ESM; in a second step I demonstrate that the application of the inference rules of TKU to tautologies of ESM will again result in tautologies of ESM.

Proving completeness is somewhat more complicated. In particular, I proceed as follows:

1. For any given wff $E\in\Lambda^k$ with $k\geq 0$ I construct - by an iterative procedure - a specific wff $F\in\Lambda^0$ with

$$F:=A_1\Rightarrow(A_2\Rightarrow(...A_M\Rightarrow E^*)...)\qquad(1)$$

such that F is a tautology of the propositional calculus if E is a tautology of ESM.

2. For $F \in \Lambda^0$ I then construct a specific wff $F^* \in \Lambda^k$ with

$$F^* := A_1^* \Rightarrow (A_2^* \Rightarrow (...A_M^* \Rightarrow E)\,...)$$

such that F^* is a theorem of TKU if F is a theorem of the propositional calculus.

3. By construction, the antecedents A_1^*,..., A_M^* are theorems of TKU, so that - by modus ponens - E is a theorem of TKU if F^* is a theorem of TKU.

4. Finally, recall (cf. Church, 1996) that a wff F is a theorem of the propositional calculus if and only if F is a tautology of the propositional calculus. Hence, combining the above steps we can conclude: if E is a tautology of ESM, then E is also a theorem of TKU.

Observe that the proof of completeness is constructive in the sense that it offers an effective way for deriving any tautology of ESM as a theorem of TKU. The following example is based on the procedure used in my completeness proof.

Example: Suppose we want to show directly that the following tautology of ESM

$$E := \neg K\,(\neg\,(p \Rightarrow p)) \in \Lambda^1$$

is also a theorem of TKU. At first construct the wff

$$F := (q \Rightarrow \neg\,(p \Rightarrow p)) \Rightarrow \neg q \in \Lambda^0$$

and observe that F is a tautology of the propositional calculus, implying that F is a theorem of TKU. Thus, we can apply the substitution rule to F in order to obtain the following theorem of TKU

$$F^* := (K\,(\neg\,(p \Rightarrow p)) \Rightarrow \neg\,(p \Rightarrow p)) \Rightarrow \neg K\,(\neg\,(p \Rightarrow p)) \in \Lambda^1$$

where q is replaced by $K\,(\neg\,(p \Rightarrow p))$. By an application of the substitution rule to the knowledge axiom K, the antecedent

$$K\,(\neg\,(p \Rightarrow p)) \Rightarrow \neg\,(p \Rightarrow p)$$

in F^* is a theorem of TKU so that we derive, by modus ponens,

$$\neg K\,(\neg\,(p \Rightarrow p))$$

as a theorem of TKU.□

7 Illustration: Iterative solution concepts

It is a standard argument that so-called iterative solution concepts for strategic games (cf., e.g., Bernheim, 1984; Pearce, 1984; Tan and Werlang, 1988), such as *rationalizability* or *iterated elimination of strictly dominated strategies*, are justifiable by the two common knowledge assumptions that

(i) the structure of the game (players, strategy space, utility-payoffs) is common knowledge among the players, and

(ii) it is common knowledge among the players that no "unreasonable" strategies will be chosen.

As it turns out, however, this assertion holds only true for epistemic models where the agents reflect on the corresponding reasoning processes whereas it is not valid for the epistemic model developed in this paper. The following example illustrates this fact.

Example. Let

$$p := \text{"Strategy } s_1 \text{ is strictly dominated."}$$
$$q := \text{"Player 1 does not play strategy } s_1\text{."}$$

The common knowledge assumption (i) implies that, e.g., the following epistemic statements are true

$$K_2(p) := \text{"Player 2 knows that the strategy } s_1 \text{ is strictly dominated."}$$
$$K_1K_2(p) := \text{"Player 1 knows that player 2 knows that strategy } s_1 \text{ is strictly dominated."}$$

Accordingly, the following epistemic statements hold by common knowledge assumption (ii)

$$K_2(p \Rightarrow q) := \text{"Player 2 knows that player 1 does not play strategy } s_1 \text{ if } s_1 \text{ is strictly dominated."}$$
$$K_1K_2(p \Rightarrow q) := \text{"Player 1 knows that player 2 knows that player 1 does not play strategy } s_1 \text{ if } s_1 \text{ is strictly dominated."}$$

The standard argument for the epistemic justification of iterated solution concepts now claims that player 1 deduces from these assumptions that "Player 2 knows that player 1 does not play strategy s_1." More specifically, it is claimed that the statement $K_1K_2(q)$ must be true if the statements $K_1K_2(p)$ and $K_1K_2(p \Rightarrow q)$ are true, or equivalently, that the wff

$$K_1K_2(p \Rightarrow q) \Rightarrow (K_1K_2(p) \Rightarrow K_1K_2(q)) \tag{2}$$

is a theorem of the underlying epistemic logic.

Suppose now that TKU is extended to the multi-agent case so that the knowledge axiom K becomes

$$K_i (p' \Rightarrow q') \Rightarrow (K_i (p') \Rightarrow K_i (q')) \tag{3}$$

for $i \in \{1, 2, ...\}$. By the substitution rule, we then obtain the following theorem

$$K_1 (K_2 (p \Rightarrow q) \Rightarrow (K_2 (p) \Rightarrow K_2 (q))) \Rightarrow (K_1 (K_2 (p \Rightarrow q)) \Rightarrow K_1 (K_2 (p) \Rightarrow K_2 (q))) \tag{4}$$

whereby $K_2 (p \Rightarrow q)$ replaces p' and $K_2 (p) \Rightarrow K_2 (q)$ replaces q' in (3). Furthermore, it can be shown that the following wff is a theorem of TKU[3]

$$((K_1 (K_2 (p \Rightarrow q)) \Rightarrow K_1 (K_2 (p) \Rightarrow K_2 (q)))) \Rightarrow (K_1 K_2 (p \Rightarrow q) \Rightarrow (K_1 K_2 (p) \Rightarrow K_1 K_2 (q))) . \tag{5}$$

Assume for the moment that necessitation was a valid inference rule so that

$$K_1 (K_2 (p \Rightarrow q) \Rightarrow (K_2 (p) \Rightarrow K_2 (q)))$$

becomes a theorem, i.e., player 1 knows that player 2's reasoning process obeys the axiom (3). In that case, we can apply modus ponens to (4) in order to obtain the theorem

$$(K_1 (K_2 (p \Rightarrow q)) \Rightarrow K_1 (K_2 (p) \Rightarrow K_2 (q))) ,$$

which is the antecedent in (5) so that we can - again by modus ponens - derive (2).

That is, if necessitation was added as a valid inference rule to TKU, then the player 1 could indeed deduce that player 2 knows he will not play strategy s_1. However, it can be shown that (2) is not a tautology of ESM and therefore it is not a theorem of TKU. As a consequence, the above common knowledge assumptions (i) and (ii) are not sufficient to justify iterative solution concepts when the underlying epistemic logic is TKU rather than existing epistemic logics.

8 Concluding remarks and outlook

If an epistemic logic has a semantic interpretation by some possible world semantic of the Kripke type, each theorem of the logic must be true in all possible worlds. According to the interpretation of the knowledge operator by the possible world semantics, all theorems must hence be known to the agent. As a consequence, any epistemic logic which has a semantic interpretation by some standard model must include necessitation as a valid inference rule. In this paper I have introduced the epistemic logic TKU for

[3] A formal proof can be otained upon request from the author.

which necessitation is not a valid inference rule. While TKU therefore does not allow for a semantic interpretation by possible world semantics, I demonstrate that TKU has a semantic interpretation by the semantic model ESM which is effectively given as a iterative procedure for developing truth-tables that take account of epistemic operators. That is, every theorem of TKU is a tautology of ESM. Furthermore, I have shown that every tautology of ESM is also a theorem of TKU. This determination theorem, which links my semantic with my syntactic approach and vice versa, is this paper's main result.

My approach is clearly an extreme concept since it describes the benchmark case of a boundedly rational agent who is opposite to the agent with absolute power of reflection. My present approach therefore leaves severalrelevant issues unresolved which I would like to address in subsequent research. In what follows I sketch two such issues.[4]

In the present paper I have adopted the standard approach of modelling the relationship between the knowledge operator K and the unawareness operator U. One might alternatively ask, why - instead of abandoning necessitation at all - I have not chosen to weaken necessitation in the sense that the agent knows all theorems he is aware of. There are two ways for accomplishing this. Suppose at first I added

$$\neg U(p) \Rightarrow K(p) \tag{6}$$

as a new axiom U4 to TKU. For this new epistemic logic TKU+U4 the agent would know every statement he is aware of, implying - by the substitution rule - that he knows every theorem he is aware of. The corresponding semantic interpretation by truth-tables would then require the deletion of any rows where $U(A)$ and $K(A)$ are simultaneously false to the effect that "awareness" $\neg U(A)$ and "knowlegde" $K(A)$ statements become logically equivalent in this new model[5]. This approach would therefore be at odds with the existing epistemic literature (cf. Section 2.3 of this paper) on unawareness, which argues that unawareness should be distinguished from the absence of knowledge.

Alternatively, I could add the following inference rule to TKU

$$\frac{A}{\neg U(A) \Rightarrow K(A)} \tag{7}$$

thereby ensuring that the logical equivalence between knowledge and awareness applies only to theorems but not to arbitrary statements. That is, $\neg U(A) \Rightarrow K(A)$ is a theorem of this logic if and only if A is a theorem. In that case, however, the question arises why the agent should apply different rules of reasoning to statements which turn out

[4]I am grateful to two anonymous referees for pointing me to these issues.

[5]Observe that the converse to U4, $K(p) \Rightarrow \neg U(p)$, stating that the agent is aware of statement p if he knows p, is already a theorem of TKU.

to be theorems than to statements which are no theorems. Nevertheless, I consider it as necessary to have - in future research - a closer look at the relationship between unawareness- and knowledge statements in the case that necessitation is abandoned as a valid inference rule.

While I think that TKU has a more realistic appeal than the existing epistemic logics, which assume necessitation as a valid inference rule, the realistic appeal of TKU is nevertheless limited. One might, for instance, convincingly argue that any real-life agent knows at least the theorem $(p \Rightarrow p)$ so that the epistemic statement $K(p \Rightarrow p)$ should also be a theorem of a realistic epistemic logic. Similarly, one might argue that there exist wff whose logical equivalence is so obvious that a real-life agent should know this equivalence. Assume, e.g., that the logical equivalence between the statements p and $\neg\neg p$ is obvious to most real-life agents. Then a realistic epistemic logic should include the following wff as theorems

$$\begin{aligned} K(p) &\Rightarrow K(\neg\neg p), \\ K(\neg\neg p) &\Rightarrow K(p). \end{aligned}$$

It is straightforward to show that these wff are not theorems of TKU but could be derived as theorems if necessitation was added as a valid inference rule to TKU. Thus, dropping necessitation as an inference rule has the consequence that many - seemingly - appealing theorems of existing epistemic logics are no longer available in TKU. To quote an anonymous referee:

> "It is true and natural that if one drops necessitation, an agent might not understand the equivalence of two logically equivalent formulas. However, there are some natural equivalences, where one feels that even a very boundedly rational agent should understand the equivalence of those formulas [...]"

While I agree in principle with this point of view, there is the practical problem of judging which logical equivalence relations are obvious - and therefore known to real-life agents - and which are not. If we had a list of such candidates of obvious logical equivalences, we could easily construct a more "realistic" epistemic logic from TKU by simply adding the corresponding epistemic statements as new axioms to the axiom system of TKU. To generate such a list is ultimativly an empirical question, which, while being beyond the scope of this paper, has to be addressed if we are interested in an epistemic logic which is a realistic model of human reasoning processes.

9 Appendix

9.1 Proving soundness

Soundness is proved by combining lemma 1 and 2.

Lemma 1. *Every axiom of* TKU *is a tautology of* ESM.

Proof: Note that the axioms A1-A3 of P2 are tautologies of the propositional logic (see Church, 1996) and therefore of ESM. In Section 5 I have already demonstrated that the axioms U1 and U2 are tautologies of ESM. I leave it to the reader to verify that axiom U3 is, by step 6, also a tautology of ESM. It remains to be demonstrated that the axioms T and K are tautologies of ESM:

- $K(p) \Rightarrow p$ is a tautology of ESM.

1. According to step 1, assign all truth value combinations to the atomic sentence p:

K	(p)	$\Rightarrow$	p
	v		v
	f		f

2. According to step 3, duplicate the rows for which p is assigned v and assign v, respectively f, in the duplicated rows to $K(p)$:

K	(p)	$\Rightarrow$	p
v	v		v
f	v		v
f	f		f

3. By step 9 repeat the procedure to assign truth values to the main operator according to the rules in step 2:

K	(p)	$\Rightarrow$	p
v	v	**v**	v
f	v	**v**	v
f	f	**v**	f

In all rows the main-operator is assigned *true*, which proves the claim.□

- $K(p \Rightarrow q) \Rightarrow (K(p) \Rightarrow K(q))$ is a tautology of ESM.

1. According to step 1, assign all truth value combinations to the atomic sentences p, q:

K	$(p$	$\Rightarrow$	$q)$	$\Rightarrow$	$(K$	(p)	$\Rightarrow$	K	$(q))$
	v		v			v			v
	v		f			v			f
	f		v			f			v
	f		f			f			f

2. According to step 2, assign truth values to the *wff* $p \Rightarrow q$:

K	$(p$	$\Rightarrow$	$q)$	$\Rightarrow$	$(K$	(p)	$\Rightarrow$	K	$(q))$
	v	v	v			v			v
	v	f	f			v			f
	f	v	v			f			v
	f	v	f			f			f

3. According to step 3, duplicate the rows in which p is assigned v and assign v, respectively f, in these duplicated rows to $K(p)$. Furthermore, assign f to $K(p)$ in the rows where p is assigned f.

K	$(p$	$\Rightarrow$	$q)$	$\Rightarrow$	$(K$	(p)	$\Rightarrow$	K	$(q))$
	v	v	v		v	v			v
	v	v	v		f	v			v
	v	f	f		v	v			f
	v	f	f		f	v			f
	f	v	v		f	f			v
	f	v	f		f	f			f

4. According to step 3, duplicate the rows in which q is assigned v and assign v, respectively f, in the duplicated rows to $K(q)$. Furthermore, assign f to $K(q)$ in the rows where q is assigned f.

K	$(p$	$\Rightarrow$	$q)$	$\Rightarrow$	$(K$	(p)	$\Rightarrow$	K	$(q))$
	v	v	v		v	v		v	v
	v	v	v		v	v		f	v
	v	v	v		f	v		v	v
	v	v	v		f	v		f	v
	v	f	f		v	v		f	f
	v	f	f		f	v		f	f
	f	v	v		f	f		v	v
	f	v	v		f	f		f	v
	f	v	f		f	f		f	f

5. According to step 3, duplicate the rows for which $(p \Rightarrow q)$ is assigned v and assign v, respectively f, in the duplicated rows to $K\,(p \Rightarrow q)$. Moreover, assign f to $K\,(p \Rightarrow q)$ in rows where $(p \Rightarrow q)$ is assigned f.

K	$(p$	$\Rightarrow$	$q)$	$\Rightarrow$	$(K$	(p)	$\Rightarrow$	K	$(q))$
v	v	v	v		v	v		v	v
f	v	v	v		v	v		v	v
v	v	v	v		v	v		f	v
f	v	v	v		v	v		f	v
v	v	v	v		f	v		v	v
f	v	v	v		f	v		v	v
v	v	v	v		f	v		f	v
f	v	v	v		f	v		f	v
f	v	f	f		v	v		f	f
f	v	f	f		f	v		f	f
v	f	v	v		f	f		v	v
f	f	v	v		f	f		v	v
v	f	v	v		f	f		f	v
f	f	v	v		f	f		f	v
v	f	v	f		f	f		f	f
f	f	v	f		f	f		f	f

6. According to step 4, delete all rows for which $K\,(p \Rightarrow q)$ and $K\,(p)$ are assigned v while $K\,(q)$ is assigned f:

K	$(p$	$\Rightarrow$	$q)$	$\Rightarrow$	$(K$	(p)	$\Rightarrow$	K	$(q))$
v	v	v	v		v	v		v	v
f	v	v	v		v	v		v	v
f	v	v	v		v	v		f	v
v	v	v	v		f	v		v	v
f	v	v	v		f	v		v	v
v	v	v	v		f	v		f	v
f	v	v	v		f	v		f	v
f	v	f	f		v	v		f	f
f	v	f	f		f	v		f	f
v	f	v	v		f	f		v	v
f	f	v	v		f	f		v	v
v	f	v	v		f	f		f	v
f	f	v	v		f	f		f	v
v	f	v	f		f	f		f	f
f	f	v	f		f	f		f	f

7.and 8. Assign truth values according to step 2:

K	$(p$	$\Rightarrow$	$q)$	$\Rightarrow$	$(K$	(p)	$\Rightarrow$	K	$(q))$
v	v	v	v	**v**	v	v	v	v	v
f	v	v	v	**v**	v	v	v	v	v
f	v	v	v	**v**	v	v	f	f	v
v	v	v	v	**v**	f	v	v	v	v
f	v	v	v	**v**	f	v	v	v	v
v	v	v	v	**v**	f	v	v	f	v
f	v	v	v	**v**	f	v	v	f	v
f	v	f	f	**v**	v	v	f	f	f
f	v	f	f	**v**	f	v	v	f	f
v	f	v	v	**v**	f	f	v	v	v
f	f	v	v	**v**	f	f	v	v	v
v	f	v	v	**v**	f	f	v	f	v
f	f	v	v	**v**	f	f	v	f	v
v	f	v	f	**v**	f	f	v	f	f
f	f	v	f	**v**	f	f	v	f	f

In all rows the main-operator is assigned *true*, which proves the claim.□

Lemma 2. *If wff A is proved in* TKU *by applying the inference rules* modus ponens, simultaneous substitution, *UR1, or UR2 to tautologies of* TKU, *then A is a tautology of ESM.*

Proof:

(i) *Modus ponens.* Suppose to the contrary that there is some row in the truth table of B such that B is false while A and $(A \Rightarrow B)$ are true at all rows of their truth tables. Notice that this can only happen if (i) a row where B is false was deleted in $(A \Rightarrow B)$ according to some step 4, 6 or 7 or (ii) there exists some row in $(A \Rightarrow B)$ where A is false such that this row has been deleted in A according to some step 4, 6 or 7. Observe that case (ii) is impossible: if some row was deleted in A it cannot reappear in $(A \Rightarrow B)$. Let us therefore focus on case (i) and suppose that there exists some row

$$\begin{array}{ccc} A & \Rightarrow & B \\ \dots & & \dots \\ \mathrm{v} & \mathbf{f} & \mathrm{f} \\ \dots & & \dots \end{array}.$$

which is deleted in $(A \Rightarrow B)$ according to some step 4, 6 or 7 but not in B, i.e., B remains false at this row. The proof proceeds by demonstrating that the existence of such a row implies the existence of another row which takes on identical truth values in $(A \Rightarrow B)$ but which is not deleted in $(A \Rightarrow B)$. Thus, case (i) implies the existence of two rows

$$\begin{array}{ccc} A & \Rightarrow & B \\ \dots & & \dots \\ \mathrm{v} & \mathbf{f} & \mathrm{f} \\ \mathrm{v} & \mathbf{f} & \mathrm{f} \\ \dots & & \dots \end{array}$$

such that only one row is deleted. The other row then contradicts the assumption that $(A \Rightarrow B)$ was a tautology. I prove this claim for the deletion of rows according to step 6 by presuming that the wff $U(C)$ and $K(C)$ appear together in $(A \Rightarrow B)$ but not in A or in B, (the proofs for step 4 and 7 are similar and left to the reader).

Consider at first the case where $U(C)$ occurs in A while $K(C)$ occurs in B such that there is a row

$$\begin{array}{ccccccccc} \dots & U & (C) & \dots & \Rightarrow & \dots & K & (C) & \dots \\ & \dots & & & & & \dots & & \\ & \mathrm{v} & \mathrm{v} & & \mathbf{f} & & \mathrm{v} & \mathrm{v} & \\ & \dots & & & & & \dots & & \end{array}$$

which has to be deleted by step 6. But then, duplication by step 5, also implies the existence of a row

$$\begin{array}{ccccccccc} \ldots & U & (C) & \ldots & \Rightarrow & \ldots & K & (C) & \ldots \\ & \ldots & & & & & \ldots & & \\ & \text{f} & \text{v} & & & & \text{v} & \text{v} & \\ & \ldots & & & & & \ldots & & \end{array}$$

which only differs in the truth value at $U(C)$. As a consequence, the truth value of B remains false whereas A is, by assumption, true. Thus, we have for this row

$$\begin{array}{ccccccccc} \ldots & U & (C) & \ldots & \Rightarrow & \ldots & K & (C) & \ldots \\ & \ldots & & & & & \ldots & & \\ & \text{f} & \text{v} & & \text{f} & & \text{v} & \text{v} & \\ & \ldots & & & & & \ldots & & \end{array}$$

, i.e., $(A \Rightarrow B)$ becomes false. Finally, observe that this row is not deleted since $U(C)$ is false.

Consider now the case where $K(C)$ occurs in A while $U(C)$ occurs in B such that there is a row

$$\begin{array}{ccccccccc} \ldots & K & (C) & \ldots & \Rightarrow & \ldots & U & (C) & \ldots \\ & \ldots & & & & & \ldots & & \\ & \text{v} & \text{v} & & \mathbf{f} & & \text{v} & \text{v} & \\ & \ldots & & & & & \ldots & & \end{array}$$

which has to be deleted by step 6. By duplication according to step 3, there must exist a row

$$\begin{array}{ccccccccc} \ldots & K & (C) & \ldots & \Rightarrow & \ldots & U & (C) & \ldots \\ & \ldots & & & & & \ldots & & \\ & \text{f} & \text{v} & & \mathbf{f} & & \text{v} & \text{v} & \\ & \ldots & & & & & \ldots & & \end{array}$$

such that the truth value of B remains false whereas A is, by assumption, true. Since $K(C)$ is here false this row is not deleted implying that $(A \Rightarrow B)$ cannot be a tautology.$\square$

(ii) *Simultaneous substitution.* If every occurrence of an atomic sentence p is replaced in A by some *wff* B, i.e., $A\{p/B\}$, then the truth table of $A\{p/B\}$ has - after duplication and deletion according to steps 3-7 - at least as many rows as the truth table of A (compare the reasoning under (i) modus ponens). For the additional rows the truth value of B is - analogously to p - either true or false so that $A\{p/B\}$ remains a tautology if A is a tautology.□

(iii) *UR1.* Suppose to the contrary that the application of *UR1* results in a *wff*

$$U(A) \Rightarrow \neg K(\neg K(...\neg K(A)))$$

that is not a tautology. That is, $U(A)$ is assigned true while $\neg K(\neg K(...\neg K(A)))$ is assigned false, which is only possible at rows where $K(\neg K(...\neg K(A)))$ is also assigned true. But according to step 7 all rows are deleted such that $U(A)$ and $K(\neg K(....\neg K(A)))$ are simultaneously true.□

(iv) *UR2.* Suppose to the contrary that the application of *UR2* results in a *wff*

$$U(A) \Rightarrow \neg K(\neg K(...\neg K(U(A))))$$

that is not a tautology. That is, $U(A)$ and $K(\neg K(...\neg K(U(A))))$ must be simultaneously true, which is impossible by step 6.□□

9.2 Proving completeness

At first I describe an iterative procedure by which I construct for any given $E \in \Lambda^k$, $k \geq 0$, the corresponding wff (1). If $E \in \Lambda^0$ simply set $F := E$ in order to derive (1). Consider now some wff $E \in \Lambda^k$ such that $k \geq 1$.

At the first stage of the iterative procedure replace in E all occurrences of the separated epistemic wff

$$K(B_1), K(B_2), ..., U(B_n), U(B_{n+1}), ...$$

by the following variables for atomic sentences:

$$p_{K(B_1)}, p_{K(B_2)}, ..., p_{U(B_n)}, p_{U(B_{n+1})}, ...$$

and denote the resulting wff as E^* whereby $E^* \in \Lambda^0$. Now take E^* as consequent and successively add in several steps antecedents $A^1_{j_i}$, $i = 1, 2, ...$, in order to obtain the following wff

$$F^1 := A^1_{j_7} \Rightarrow ... \left(A^1_{j_6} \Rightarrow ... \left(A^1_{j_5} \Rightarrow ... \left(A^1_{j_4} \Rightarrow ... \left(A^1_{j_3} \Rightarrow ... \left(A^1_{j_2} \Rightarrow ... \left(A^1_{j_1} \Rightarrow ...E^*\right)\right)\right)\right)\right)\right).$$

1. For all $p_{K\left(B_{j_1}\right)}$ add the following antecedents to the consequent E^*:

$$A^1_{j_1} := \left(p_{K\left(B_{j_1}\right)} \Rightarrow B_{j_1}\right)$$

2. For all $p_{K\left(B_{j_2}\right)}$ such that $B_{j_2} = U\left(D_{j_2}\right)$ for some wff D_{j_2} add the following antecedents to the wff resulting from step 1:

$$A^1_{j_2} := \left(U\left(D_{j_2}\right) \Rightarrow \neg p_{K\left(B_{j_2}\right)}\right)$$

3. For all $p_{K\left(B_{j_3}\right)}$ such that $B_{j_3} = \neg K\left(...\neg K\left(U\left(D_{j_3}\right)\right)\right)$ for some wff D_{j_3} add the following antecedents to the wff resulting from step 2:

$$A^1_{j_3} := \left(U\left(D_{j_3}\right) \Rightarrow \neg p_{\neg K\left(...\neg K\left(U\left(D_{j_3}\right)\right)\right)}\right)$$

4. For all $p_{K\left(B'_{j_4}\right)}$ and $p_{K\left(B_{j_4}\right)}$ such that $B_{j_4} = \left(B'_{j_4} \Rightarrow B''_{j_4}\right)$ for some wff B'_{j_4}, B''_{j_4} add the following antecedents to the wff resulting from step 3:

$$A^1_{j_4} := \left(p_{K\left(B_{j_4}\right)} \Rightarrow \left(p_{K\left(B'_{j_4}\right)} \Rightarrow p_{K\left(B''_{j_4}\right)}\right)\right)$$

5. For all $p_{U\left(B_{j_5}\right)}$ add the following antecedents to the wff resulting from step 4:

$$A^1_{j_5} := \left(p_{U\left(B_{j_5}\right)} \Rightarrow \left(B_{j_5} \Rightarrow B_{j_5}\right)\right)$$

6. For all $p_{U\left(B_{j_6}\right)}$ and $p_{K\left(B_{j_6}\right)}$ add the following antecedents to the wff resulting from step 5:

$$A^1_{j_6} := \left(p_{U\left(B_{j_6}\right)} \Rightarrow \neg K\left(B_{j_6}\right)\right)$$

7. For all $p_{U\left(B_{j_7}\right)}$ and $p_{K(B)}$ such that $B = \neg K\left(...\neg K\left(B_{j_7}\right)\right)$ add the following antecedents to the wff resulting from step 6:

$$A^1_{j_7} := \left(p_{U\left(B_{j_7}\right)} \Rightarrow \neg K\left(...\neg K\left(B_{j_7}\right)\right)\right)$$

Observe that $F^1 \in \Lambda^{k-1}$: there do not occur any epistemic wff in E^*; the antecedents $A^1_{j_1}$, respectively $A^1_{j_5}$, contain wff B whose epistemic degree is exactly one less than that of the wff $K(B)$, respectively $U(B)$; moreover, the remaining antecedents have a strictly smaller epistemic degree than E.

In case $k = 1$, set $F := F^1 \in \Lambda^0$ and observe that we are done deriving the wff (1). If instead $k > 1$, then proceed to the second stage of the iterative procedure and replace in F^1 all occurrences of the separated epistemic wff

$$K\left(C_1\right), K\left(C_2\right), ..., U\left(C_n\right), U\left(C_{n+1}\right), ...$$

by the variables:

$$p_{K(C_1)}, p_{K(C_2)}, ..., p_{U(C_n)}, p_{U(C_{n+1})}, ...$$

Denote the resulting wff as F^{1*} whereby $F^{1*} \in \Lambda^0$. As in the first stage of the iterative procedure, construct now a wff

$$F^2 := A^2_{j7} \Rightarrow ... \left(A^2_{j6} \Rightarrow ... \left(A^2_{j5} \Rightarrow ... \left(A^2_{j4} \Rightarrow ... \left(A^2_{j3} \Rightarrow ... \left(A^2_{j2} \Rightarrow ... \left(A^2_{j1} \Rightarrow ... F^{1*}\right)\right)\right)\right)\right)\right)$$

whereby the antecedents A^2_i are added according to the above steps 1-7. Notice that $F^2 \in \Lambda^{k-2}$. Thus, if $k = 2$ set $F := F^2 \in \Lambda^0$ in order to obtain the desired wff (1). If instead $k > 2$ we have to go through another stage of the iterative procedure in order to construct a wff $F^3 \in \Lambda^{k-3}$ in the obvious way. For $E \in \Lambda^k$ we arrive after exactly k stages at the corresponding wff (1).

Lemma 3: *Consider some wff $E \in \Lambda^k$, $k \geq 1$, and the corresponding wff (1), which had been constructed by the iterative procedure described above. If E is a tautology of ESM then F is a tautology of the propositional calculus.*

Proof: While E is, by assumption, a tautology, the same is not necessarily true for the consequent E^* in (1). Since all epistemic wff in E are replaced in E^* by variables for atomic propositions, there may exist some row in the truth table of E^* that had been deleted in the development of the truth table of E. If E^* is false at such a row it would not be a tautology despite the fact that E is a tautology. The proof of lemma 3 proceeds by demonstrating that whenever the consequent E^* may be false at some row, some antecedent A_i, $i = 1, 2, ...$, in (1) must also be false. As a consequence, F would be true at such rows thereby proving that F is a tautology of the propositional calculus.

At first observe that the iterative procedure for generating (1) ensures that there occur $p_{K(B)}$, $p_{U(B')}$ in F for any epistemic wff $K(B)$, $U(B')$ occurring in E. Because of the steps 3,4,6,7 of the iterative procedure of the development of truth-tables there may exist truth-value combinations for E^* that do not appear in the truth table of E. In what follows I comprehensively investigate all possible cases.

Case 1: According to step 3 we have for E rows such that

B	$K(B)$
v	v
v	f
f	f

whereas we have for E^*

B	$p_{K(B)}$
v	v
v	f
f	v
f	f

That is, the truth tables of E^* and E differ because there are rows in E^* where B is false whereas $p_{K(B)}$ is true. For such rows, however, the antecedent

$$\left(p_{K(B)} \Rightarrow B\right),$$

resulting from step 1 of the iterative procedure for generating F, must be false. Thus, F is true at such rows.□

Case 2: According to step 4 we have for E rows

$K(B' \Rightarrow B'')$	$K(B')$	$K(B'')$
v	v	v
v	f	v
v	f	f
f	v	v
f	v	f
f	f	v
f	f	f

whereas we have for E^*

$p_{K(B' \Rightarrow B'')}$	$p_{K(B')}$	$p_{K(B'')}$
v	v	v
v	v	f
v	f	v
v	f	f
f	v	v
f	v	f
f	f	v
f	f	f

Thus, the truth tables of E^* and E differ because there are rows in E^* where $p_{K(B' \Rightarrow B'')}$ and $p_{K(B')}$ are true while $p_{K(B'')}$ is false. By step 4 of the iterative procedure for generating F, there exists some antecedent

$$\left(p_{K(B' \Rightarrow B'')} \Rightarrow \left(p_{K(B')} \Rightarrow p_{K(B'')}\right)\right)$$

which is false at such rows.□

Case 3: By step 6, we have for E rows such that

$U(B)$	$K(U(B))$
v	f
f	v
f	f

whereas we have for E^*

$p_{U(B)}$	$p_{K(U(B))}$
v	v
v	f
f	v
f	f

In that case the truth tables of E^* and E differ because there are rows in E^* where $p_{U(B)}$ and $p_{K(U(B))}$ are simultaneously true. For such rows the antecedent

$$\left(U(B) \Rightarrow \neg p_{K(U(B))}\right),$$

resulting from step 2 of generating F, ensures that F is true.□

Case 4: Also by step 6, we have for E rows such that

$U(B)$	$K(\neg K(....\neg K(U(B))))$
v	f
f	v
f	f

whereas we have for E^*

$p_{U(B)}$	$p_{K(\neg K(....\neg K(U(B))))}$
v	v
v	f
f	v
f	f

Here the antecedent

$$\left(U\left(B\right) \Rightarrow \neg p_{K(\neg K(....\neg K(U(B))))}\right),$$

resulting from step 3 of generating F, guarantees that F is true.□

We leave it to the reader to verify in a similar way that the steps 6 and 7 of the iterative procedure of generating wff F ensure that F is true whenever the rows of E and E^* differ because of step 7 of the iterative procedure for the development of truth-tables. This finally proves the lemma 3.□□

Lemma 4: *Consider some wff F which had been constructed for some wff $E \in \Lambda^k$, $k \geq 1$, by the iterative procedure described above. Denote by F^* the wff that results when all occurrences of variables for atomic propositions $p_{K(B)}$, $p_{U(B')}$ in F are replaced by the corresponding wff $K\left(B\right)$, $U\left(B'\right)$, i.e.,*

$$F^* := A_1^* \Rightarrow \left(A_2^* \Rightarrow \left(...A_M^* \Rightarrow E\right)...\right) \tag{8}$$

If F is a theorem of the propositional calculus, then E is a theorem of TKU.

Proof: Observe at first that F^* is - by the substitution rule - a theorem of TKU if F is a theorem of the propositional calculus. Thus, if we can establish that all antecedents A_i^*, $i = 1, 2, ...$, in (8) are theorems of TKU, the E is, by repeated application of modus ponens, a theorem of TKU. According to the seven steps of the iterative procedure of generating the wff (1) there are seven different cases of how the antecedents A_i^*, $i = 1, 2, ...$, may look like. In the remainder of the proof I demonstrate for each case that the corresponding antecedent is a theorem of TKU.

Case 1. According to step 1 we have

$$A_{j_1} := \left(p_{K\left(B_{j_1}\right)} \Rightarrow B_{j_1}\right)$$

so that the corresponding antecedent becomes

$$A_{j_1}^* := \left(K\left(B_{j_1}\right) \Rightarrow B_{j_1}\right).$$

Applying the substitution rule to the knowledge axiom K shows that $A_{j_1}^*$ is a theorem.□

Case 2. By step 2,

$$A_{j_2} := \left(U\left(D_{j_2}\right) \Rightarrow \neg p_{K\left(U\left(D_{j_2}\right)\right)}\right)$$

so that

$$A^*_{j_2} := \left(U\left(D_{j_2}\right) \Rightarrow \neg K\left(U\left(D_{j_2}\right)\right)\right),$$

which is a theorem of TKU by axiom U3 and the substitution rule.□

Case 3. By step 3,

$$A_{j_3} := \left(U\left(D_{j_3}\right) \Rightarrow \neg p_{\neg K\left(\ldots\neg K\left(U\left(D_{j_3}\right)\right)\right)}\right)$$

so that

$$A^*_{j_3} := \left(U\left(D_{j_3}\right) \Rightarrow \neg K\left(\ldots\neg K\left(U\left(D_{j_3}\right)\right)\right)\right).$$

Since $U(p) \Rightarrow \neg K(U(p))$ is an axiom of TKU, a repeated application of the inference rule UR2 and an application of the substitution rule establishes $A^*_{j_3}$ as a theorem of TKU.□

Case 4. By step 4,

$$A_{j_4} := \left(p_{K\left(B'_{j_4} \Rightarrow B''_{j_4}\right)} \Rightarrow \left(p_{K\left(B'_{j_4}\right)} \Rightarrow p_{K\left(B''_{j_4}\right)}\right)\right)$$

so that

$$A^*_{j_4} := \left(K\left(B'_{j_4} \Rightarrow B''_{j_4}\right) \Rightarrow \left(K\left(B'_{j_4}\right) \Rightarrow K\left(B''_{j_4}\right)\right)\right),$$

which is a theorem by axiom T and the substitution rule.□

Case 5. By step 5,

$$A_{j_5} := \left(p_{U\left(B_{j_5}\right)} \Rightarrow \left(B_{j_5} \Rightarrow B_{j_5}\right)\right)$$

so that

$$A^*_{j_5} := \left(U\left(B_{j_5}\right) \Rightarrow \left(B_{j_5} \Rightarrow B_{j_5}\right)\right).$$

Notice that $q \Rightarrow (p \Rightarrow p)$ is a theorem of the propositional calculus so that, by the substitution rule, $A^*_{j_5}$ is also a theorem.□

Case 6. By step 6,

$$A_{j_6} := \left(p_{U\left(B_{j_6}\right)} \Rightarrow \neg K\left(B_{j_6}\right)\right)$$

so that

$$A^*_{j_6} := \left(U\left(B_{j_6}\right) \Rightarrow \neg K\left(B_{j_6}\right)\right),$$

which is a theorem by axiom U2 and the substitution rule.□

Case 7. By step 7,

$$A_{j_7} := \left(p_{U\left(B_{j_7}\right)} \Rightarrow \neg K\left(...\neg K\left(B_{j_7}\right)\right)\right)$$

so that

$$A_{j_7}^* := \left(U\left(B_{j_7}\right) \Rightarrow \neg K\left(...\neg K\left(B_{j_7}\right)\right)\right).$$

Since $U(p) \Rightarrow \neg K(p)$ is an axiom of TKU, a repeated application of the inference rule UR1 and an application of the substitution rule gives $A_{j_7}^*$ as a theorem of TKU.□□

References

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

Bernheim, B.D. (1984), "Rationalizable strategic behavior," *Econometrica* **52**, 1007-1028.

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

Church, A. (1996), *Introduction to mathematical logic*, Princeton University Press: Princeton.

Dekel, E., Lipman, B.L., and A. Rustichini (1998a), "Standard state-space models preclude unawareness," *Econometrica* **66**, 159-173.

Dekel, E., Lipman, B.L., and A. Rustichini (1998b), "Recent developments in modeling unforeseen contingencies," *European Economic Review* **42**, 523-542.

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

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

Halpern, J.Y., and L.C. Rego (2005), "Interactive unawareness revisited," *Theoretical Aspects of Rationality and Knowledge: Proceedings of the Tenth Conference (TARK 2005)*, 78-91.

Heifetz, A., Meier, M., and B. Schipper (2005), "Interactive unawareness," *Journal of Economic Theory*, forthcoming.

Kaneko, M., and N.Y. Suzuki (2003), "Epistemic models of shallow depths and game theoretical decision making: Horticulture," *Journal of Symbolic Logic* **68**, 163–186.

Kaneko, M., and N.Y. Suzuki (2005), "Contentwise complexity of inferences in epistemic logics of shallow depths I: General development," Unpublished manuscript.

Kripke, S.A. (1963), "Semantical analysis of modal logic I Normal modal propositional calculi," *Zeitschr. f. math. Logik und Grundlagen d. Math.* **9**, 67–96.

Modica, S. and A. Rustichini (1994), "Unawareness: A formal theory of unforeseen contingencies," *Theory and Decision* **37**, 107—124.

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

Pearce, D.G. (1984), "Rationalizable strategic behavior and the problem of perfection," *Econometrica* **52**, 1029-1050.

Tan, T., and S.R. Costa Da Werlang (1988), "On Aumann's notion of common knowledge, an alternative approach," *Journal of Economic Theory* **45**, 370-391.

Wittgenstein, L. (1922), *Tractatus logico-philosophicus*, reprinted in Ludwig Wittgenstein *Tractatus logico-philosophicus, Philosophische Untersuchungen* (1990), [ed.] P. Philipp, Reclam Verlag: Leipzig.