

# “Knowing Whether,” Meta-Knowledge, and Epistemic Bounded Rationality\*

Konrad Grabiszewski<sup>†</sup>

March 22, 2011

*To know that we know what we know, and to know that we do not know what we do not know, that is true knowledge.*

Nicolaus Copernicus

## Abstract

In economics, it is a standard assumption that an agent knows the model. We investigate the implications of this assumption, which leads an agent to conduct self-analysis, in the case of the model of knowledge. We show that an agent cannot simultaneously know the model and be epistemically boundedly rational. Our analysis relies on the idea of “knowing whether” introduced by Hart et al. [15] and expands our understanding of their operator. As a consequence of our results, we discovered that in Aumann’s Agreement Theorem it is not necessary that the agents’ possibility correspondences are commonly known. We also argue that we cannot avoid the no-trade theorems by introducing non-partitional knowledge structure while simultaneously assuming that the agents know the model.

Keywords: Knowledge; Meta-knowledge; Epistemic Bounded Rationality; Agreeing to Disagree.

JEL classification: D80; D82

---

\*I thank Adam Brandenburger and Amanda Friedenberg for discussions about knowledge and logic. I also thank Joseph Halpern, Erkut Ozbay, David Pearce, David Ross, Vasiliki Skreta and the audiences at NYU, ITAM, University of Alicante and National University of Singapore for their comments. Support from the Asociación Mexicana de Cultura A.C., and the Stern School of Business is gratefully acknowledged.

<sup>†</sup>Address: Department of Business Administration, ITAM, Rio Hondo 1, Col. Progreso Tizapan, C.P. 01080 Mexico D.F., Mexico, konrad.grabiszewski@itam.mx, <http://allman.rhon.itam.mx/~konrad.grabiszewski/>

# 1 Introduction

Economists usually (implicitly) assume that the model they construct is known to the agents about whom the model is created. While the analyst builds the model in order to describe or predict reality, the agent, Ann, who knows the model – that is, has access to the same tools as the analyst, conducts self-analysis. In this paper, we focus on the model of knowledge and Ann who knows it. We investigate the properties of her knowledge in this setup.

The standard model of knowledge consists of a state space,  $\Omega$ , and a possibility correspondence,  $P$ , defined on it. There are three knowledge properties that interests us: (1) Truth Axiom (e.g., if Ann knows that it rains, then it actually rains); (2) Positive Introspection Axiom (e.g., if Ann knows that it is cold, then she knows that she knows it is cold); and (3) Negative Introspection Axiom (e.g., if Ann does not know that Bob is sick, then she knows that she does not know that Bob is sick).

Geanakoplos [11] and Rubinstein [22] suggest that bounded rationality be captured by violation of at least one of these properties. Since this type of bounded rationality is related to knowledge, we call it epistemic bounded rationality. We argue that Ann, who knows her own model of knowledge, must satisfy all knowledge axioms. In addition, she knows about this fact.

The problem of the agent knowing her own possibility correspondence has already been mentioned in Aumann [1]. His famous Agreement Theorem requires that agents' priors on a state space are the same and that their possibility correspondences satisfy all knowledge axioms. The theorem states that agents' posteriors must be the same if they are commonly known. However, Aumann also added “(...) the implicit assumption that the information partitions  $\mathcal{P}_1$  and  $\mathcal{P}_2$  are themselves common knowledge.” This assumption is repeated in Aumann [2]: “While Player 1 may well be ignorant of what Player 2 knows (...) 1 cannot be ignorant of the partition  $\mathcal{S}_2$  itself. (...) Indeed, since the specification of each  $\omega$  includes a complete description of the state of the world, it includes also a list of those other states  $\omega'$  of the world that are, for Player 2, indistinguishable from  $\omega$ .”

However, as pointed by, among others, Fagin et al. [10] and Gilboa [13], such a specification leads to circularity. If a state,  $\omega$ , is to contain a description of possibility correspondence,  $P$ , then we need a well-defined  $P$  before we can construct  $\Omega$ . However,  $P$  is a function defined over  $\Omega$ , so we need to construct  $\Omega$  before we can describe  $P$ . In consequence, it is not clear how to proceed with construction of a model  $(\Omega, P)$ . Hierarchical structures based on the idea of type space of Harsanyi [14] solve this issue (see, for instance, Brandenburger and Dekel [6], Mertens and Zamir [18], Heifetz [16]). As an alternative, Heifetz [17] suggests using non-well-founded set theory to solve the problem of circularity.

Aumann [3] provides alternative meaning of the “the implicit assumption that the information partitions  $\mathcal{P}_1$  and  $\mathcal{P}_2$  are themselves common knowledge.” In Section 7, Aumann addresses the “question (...) whether each individual “knows” the information partitions  $\mathcal{S}_i$  of the other.” His

answer implies that by *knowing* a partition he means, what we can informally state as, *understanding* definition of partition: “In brief,  $j$ ’s knowing the operator  $k_i$  means simply that  $j$  knows what it means for  $i$  to know something, not that  $j$  knows anything specific about what  $i$  knows. Thus the assertion that each individual “knows” the knowledge operators  $k_i$  of all individuals has no real substance; it is part of the framework.” The use of quotation marks is explained in footnote 17: “It should be recognized that “knowledge” in this connection has a meaning that is somewhat different from that embodied in the operators  $k_i$  and  $K_i$ ; that is why we have been using quotation marks when talking about “knowing” an operator or a partition. That an individual  $i$  knows an event  $E$  or a formula  $f$  can be embodied in formal statements ( $\omega \in K_i E$  or  $k_i f \in \omega$ ) that are well defined within the formal systems we have constructed, and whose truth or provability can be explicitly discussed in the context of these formal systems. This is not the case for “knowing” an operator or a partition.”

Hence, we can think of *knowing* an information structure in two ways, either in a formal sense of knowledge (i.e., Ann knowing her own possibility correspondence means that she knows what states she would consider as possible for any given state  $\omega$ ) or in an informal sense of understanding the meaning (i.e., Ann knowing her own possibility correspondence means that she understands what a possibility correspondence is)<sup>1</sup>. Since we are interested in analyzing the consequences of the agent knowing the model of knowledge that describes her knowledge, it is the former interpretation that we follow in this paper. In order to distinguish between knowledge captured in a model  $(\Omega, P)$  and knowledge of the model, the former is called *meta-knowledge*. The introduction of meta-knowledge suggests considering meta-meta-knowledge, that is Ann’s knowledge about her meta-knowledge. Next, we may think of meta-meta-meta-knowledge and, in consequence, we would obtain an infinite structure of knowledge. However, such a construction is not of our interest in this paper. We want to investigate the properties of the underlying model,  $(\Omega, P)$ , and focusing only on meta-knowledge is enough for our purposes. Naturally, our result extends to higher level of meta structure. In particular, if Ann meta-meta-knows her model of meta-knowledge, then such a model must satisfy all knowledge axioms.

In Section 2, we briefly review the standard, event-based approach to modeling knowledge that was introduced by Aumann [1]. We also discuss how Ann’s knowledge and its properties are captured by the state space and possibility correspondence,  $(\Omega, P)$ . More extensive and advanced analysis of this model is provided by, among others, Dekel and Gul [8], Fagin et al. [9], Geanakoplos [11], and Rubinstein [22]. Brandenburger [4], Geanakoplos [12], Reny [21], and Samuelson [25] survey knowledge modeling and its applications in economics. Section 3 is the first step of our analysis. We discuss the “knowing whether” operator introduced by Hart et al. [15]. We argue that assuming the Truth Axiom and that Ann knows whether the knowledge axioms hold implies that she knows that she is epistemically unboundedly rational and, in fact, is correct (Proposition 3.1). We also show

---

<sup>1</sup>In the same spirit, *knowing* a function  $f$  between  $X$  and  $Y$  might mean knowing what  $y$  is assigned to each  $x$  or just understanding the definition of function.

that replacing the Truth Axiom with either the Positive or Negative Introspection Axioms makes Ann believe that she satisfies the knowledge axiom. However, since the Truth Axiom does not have to hold, Ann does not have to know this, that is, although she assigns measure 1 to her not being epistemically boundedly rational, she might be wrong (Proposition 3.2). In Section 4, we conduct the second step of our analysis. We discuss the model of meta-knowledge,  $(X, P_X)$ , that captures Ann’s reasoning about her own model of knowledge. We build meta-state space and show that, under the assumption of Ann meta-knowing her model of knowledge, she must know whether the knowledge axioms hold in the underlying model of knowledge,  $(\Omega, P)$  (Proposition 4.1). When we combine this with the result obtained in Section 3, we infer that the agent knows that all knowledge axioms hold. In consequence, if the model of knowledge is known to the agent who is epistemically boundedly rational, then the model becomes inconsistent. Finally, in Section 5, we present two applications of our main result. First, building on Samet [23], we show why we do not need the knowledge of possibility correspondences in the Agreement Theorem of Aumann [1]. Second, we conclude that the no-trade theorem (Milgrom and Stokey [19]) cannot be avoided by violation the of knowledge axioms, if the setup of the theorem itself is known to the agents.

## 2 Modeling Knowledge and Knowledge Properties

We begin with the standard model of Ann’s knowledge, which consists of a state space,  $\Omega$ , and possibility correspondence,  $P : \Omega \rightarrow 2^\Omega$ . A **state**,  $\omega$ , is a description of the world. States are internally consistent, mutually exclusive, and exhaustive. An **event**,  $E$ , is a subset of  $\Omega$ . For example, an event “it rains” is a collection of all states at which “it rains.” We will denote the complement of  $E$  by  $\neg E$ .

Ann’s perception of the world is represented via a possibility correspondence. If  $\omega$  is a true state, then  $P(\omega)$  is a subset of  $\Omega$  that contains all states Ann considers as possible. However, it is very important to stress that we must not take the verb *considers* literally. Ann does not perceive the world in terms of states.  $\Omega$  and  $P$  are analyst’s tools. When Ann “considers”  $\omega'$  as a possible state, we think of Ann *as if* she were considering  $\omega'$  as a possible state.

There are three standard properties of knowledge that are of interest here: the Truth Axiom, the Positive Introspection Axiom and the Negative Introspection Axiom. These **knowledge axioms** are formally represented using possibility correspondence and state space.

The first property, the **Truth Axiom**, is also called non-delusion (see Geanakoplos [11]). It says that whatever Ann knows must be true. The Truth Axiom is a part of epistemic *savoir vivre*: it distinguishes between knowledge and belief. Ann *believes* that it rains if she considers “it does not rain” as impossible. That is, from a probabilistic point of view, she assigns measure 1 to the event “it rains.” If, in addition, it does rain, then we say that Ann *knows* that “it rains.” The distinction

between *knowledge* and *belief* is not observable by Ann – it is only a conceptual differentiation used by the analyst. Encoding the Truth Axiom via  $P$  implies that a true state must always be considered as possible – i.e.,  $\omega \in P(\omega)$ . The **Positive Introspection Axiom** says that whenever Ann knows something, she knows that she knows it.  $P(\omega') \subset P(\omega)$  if  $\omega' \in P(\omega)$  captures this property. Finally, the **Negative Introspection Axiom** requires that, if there is a fact Ann does not know, she knows about her own lack of knowledge.  $P$  captures this if  $P(\omega) \subset P(\omega')$  whenever  $\omega' \in P(\omega)$ .

Using the state space and possibility correspondence, agent's knowledge is formally defined in the following way: *Ann knows event  $E$  at state  $\omega$*  if  $P(\omega)$  is a subset of  $E$ . This set-theoretical definition of knowledge leads to construction of a **knowledge operator**,  $K$ , defined on subsets of  $\Omega$  in the following way:  $K(E) = \{\omega : P(\omega) \subset E\}$ .  $KE$  is an event – a collection of all states at which Ann knows  $E$ .

Every operator,  $K$ , defined from  $P$  satisfies the following properties, irrespective of the properties of  $P$ :

(N)  $K\Omega = \Omega$ . “N” stands for Necessitation; this property says that at each state Ann knows that the whole state space obtains – in other words, Ann knows that something happens.

(MC<sup>2</sup>)  $K(E \cap F) = KE \cap KF$ . In other words, Ann knows both events  $E$  and  $F$  if and only if she knows  $E$  and she knows  $F$ . It is well known that MC implies Monotonicity – if  $E \subset F$ , then  $KE \subset KF$ . As such, if Ann knows event  $E$  that is a refinement of event  $F$ , then she also knows event  $F$ .

In turn, knowledge axioms of  $P$  generate knowledge axioms of  $K$ :

(1) If the Truth Axiom holds for  $P$ , then for each state,  $\omega$ , and each event,  $E$ ,  $\omega \in \neg KE \cup E$ . In order to make this more transparent, the Truth Axiom also is expressed as  $\omega \in KE \rightarrow E$ .

(2) If the Positive Introspection Axiom holds for  $P$ , then for each state,  $\omega$ , and each event,  $E$ ,  $\omega \in \neg KE \cup KKE$ , or more transparently,  $\omega \in KE \rightarrow KKE$ .

(3) If the Negative Introspection Axiom holds for  $P$ , then for each state  $\omega$  and each event  $E$ ,  $\omega \in KE \cup K\neg KE$ , or more transparently,  $\omega \in \neg KE \rightarrow K\neg KE$ .

We say that the agent is **epistemically boundedly rational** if at least one knowledge axiom is violated.

---

<sup>2</sup>This term appears in Dekel and Gul [8]. It is derived from terminology of modal logic. MC combines two axioms, M, which says that  $K(E \cap F) \subset KE \cap KF$ , and C, which says that  $KE \cap KF \subset K(E \cap F)$  (see Chellas [7]). Samet [24] calls this property the Distribution Axiom.

### 3 “Knowing Whether” and Epistemic Bounded Rationality

Hart et al. [15] proposes a new knowledge operator,  $J$ , called *knowing whether* and defined in the following way:  $JE = KE \cup K\neg E$ . If Ann *knows whether*  $E$ , she either knows that  $E$  is true or she knows that  $E$  is not true. She is not allowed not to know – she must have an opinion.

We are interested in applying the *knowing whether* operator to knowledge axioms. Ann knows whether the Truth Axiom is true if, for each state  $\omega$  and each event,  $E$ ,  $\omega \in K(\neg KE \cup E) \cup K(KE \cap \neg E)$ . She knows whether she satisfies the Positive Introspection Axiom if, for each state  $\omega$  and each event,  $E$ ,  $\omega \in K(\neg KE \cup KKE) \cup K(KE \cap \neg KKE)$ . Finally, Ann knows whether the Negative Introspection Axiom holds if, for each state  $\omega$  and each event,  $E$ ,  $\omega \in K(KE \cup K\neg KE) \cup K(\neg KE \cap \neg K\neg KE)$ .

#### Proposition 3.1

*Assume that the agent knows whether knowledge axioms hold. If the Truth Axiom holds, then the agent knows that she satisfies all knowledge axioms.*

#### Proof:

Assuming the Truth Axiom implies that it is a tautology and, in consequence, that Ann knows the axiom is true. To prove that she knows that the Positive Introspection Axiom is true, suppose that she does not know. That is, fix some  $\omega$  and  $E$  and assume that  $\omega \in K(KE \cap \neg KKE)$ . Due to the MC property,  $K(KE \cap \neg KKE) = KKE \cap K\neg KKE$ . The Truth Axiom implies that  $K\neg KKE = \neg KKE$ . In consequence,  $\omega \in KKE \cap \neg KKE$ , which is a contradiction. Hence, it must be true that  $\omega \in K(\neg KE \cup KKE)$ . In order to see that Ann must know that she satisfies the Negative Introspection Axiom, we also rely on proof by contradiction. Suppose that  $\omega \in K(\neg KE \cap \neg K\neg KE)$ . Since  $K(\neg KE \cap \neg K\neg KE) = K\neg KE \cap K\neg K\neg KE = K\neg KE \cap \neg K\neg KE$ , we again obtain a contradiction. ■

Proposition 3.1 provides additional insight into our understanding of the *knowing whether* operator. While  $J$  is by definition more general than  $K$ , it becomes  $K$  when applied to knowledge axioms under assumption of the Truth Axiom. In other words, the Truth Axiom and Ann knowing whether she violates epistemic bounded rationality implies that she is not epistemically boundedly rational. In order to see why we need the Truth Axiom in Proposition 3.1, consider the following example in which the axiom is violated.

#### Example 3.1

*Consider the state space  $\Omega = \{\omega_1, \omega_2\}$ , with possibility correspondence defined as  $P(\omega_1) = \{\omega_2\}$  and  $P(\omega_2) = \{\omega_1\}$ . There are three possible events,  $E_1 = \{\omega_1\}$ ,  $E_2 = \{\omega_2\}$ , and  $E_3 = \Omega$ , and it is*

easy to verify that at each state and for each event, Ann knows whether the knowledge axioms hold. Note that  $\{\omega_1\} = KKE_1 = K\neg KKE_1$ . That is,  $\omega_1 \in KKE_1 \cap K\neg KKE_1 = K(KE_1 \cap \neg KKE_1)$ . In other words, at  $\omega_1$  Ann believes<sup>3</sup> that she violates the Positive Introspection Axiom for event  $E_1$ . Note also that  $\{\omega_1\} = K\neg KE_2 = K\neg K\neg KE_2$ , which implies that  $\omega_1 \in K(\neg KE_2 \cap \neg K\neg KE_2)$  – that is, Ann believes that she violates the Negative Introspection Axiom for event  $E_2$ . In both cases Ann is incorrect, as she does satisfy both introspection axioms.

It is also possible that, while Ann violates the Truth Axiom, she believes that the axiom holds.

### Example 3.2

Consider the state space from Example 3.1. Possibility correspondence is defined as  $P(\omega_1) = P(\omega_2) = \{\omega_2\}$ . We want to show that, for each event,  $E$ ,  $\omega_1 \in K(\neg KE \cup E)$ . If  $\omega_2 \in E$ , then  $P(\omega_1) \subset E$  and, in consequence,  $\omega_1 \in K(\neg KE \cup E)$ . If  $\omega_2 \notin E$ , then  $E = \{\omega_1\}$  and  $KE = \emptyset$ . Hence,  $\neg KE = \Omega$ , and again  $\omega_1 \in K(\neg KE \cup E)$ .

In Proposition 3.1, the Truth Axiom can be replaced by either the Positive or Negative Introspection Axioms. However, since we want the Truth Axiom to be violated, we also need to replace *agent knowing* with *agent believing*.

### Proposition 3.2

Assume that the agent knows whether knowledge axioms hold. If either the Positive or Negative Introspection Axiom holds, then the agent believes she satisfies all knowledge axioms.

#### Proof:

Assuming an axiom means that it is true at each state and, in consequence, that Ann knows that the axiom holds. Hence, if we assume the Positive (Negative) Introspection Axiom, then we only need to check that Ann believes that the Truth Axiom and the Negative (Positive) Introspection Axiom hold. Fix some  $\omega$  and  $E$ . First, assume the Positive Introspection Axiom. Suppose that Ann believes that she violates the Truth Axiom,  $\omega \in KKE \cap K\neg E$ . Since  $K\neg E \rightarrow KK\neg E$ , it must be true that  $\omega \in KKE \cap KK\neg E$ . But this is a contradiction, as we see if we assume that  $\omega \in KKE$  and  $\omega \in KK\neg E$ . The former implies that  $P(\omega) \subset KE$ . The latter indicates that  $P(\omega) \subset K\neg E$ . Since  $KE$  and  $K\neg E$  are disjoint sets both statements cannot be simultaneously true. Ann's belief that the Negative Introspection Axiom does not hold means that  $\omega \in K\neg KE \cap K\neg K\neg KE$ , which, in turn, implies that  $\omega \in KK\neg KE \cap K\neg K\neg KE$ . That is, Ann simultaneously believes that event  $F = K\neg KE$  holds and does not hold. Again, this is a contradiction. Next, assume the Negative

<sup>3</sup>Since the Truth Axiom is violated here, we need to replace *the agent knows* by *the agent believes*.

Introspection Axiom which says that  $\omega \in KE \cup K\neg KE$ . Suppose that Ann believes that the Truth Axiom is violated – that  $\omega \in KKE \cap K\neg E$ . By assumption, then, either  $\omega \in KE$  or  $\omega \in K\neg KE$ . The former contradicts  $\omega \in K\neg E$ , and the latter contradicts  $\omega \in KKE$ . In consequence, Ann believes that she satisfies both the Truth Axiom and the Negative Introspection Axiom. The fact that these axioms imply the Positive Introspection Axiom is a tautology and, in consequence, she believes that she satisfies the Positive Introspection Axiom as well. ■

## 4 Meta-Knowledge and Epistemic Bounded Rationality

In order to capture Ann’s reasoning about her own model of knowledge, we build a model of meta-knowledge. This model consists of meta-state space,  $X$ , and possibility correspondence,  $P_X$ , defined on  $X$ . Each meta-state,  $x$ , specifies for each pair,  $(\omega_1, \omega_2)$ , whether  $\omega_1 \in P(\omega_2)$  is true or false (but never both). We should think of formulas  $\omega_1 \in P(\omega_2)$  as being atomic sentences from which we can construct, via standard logical methods, longer strings. In other words,  $x$  represents some model of knowledge,  $(\Omega, P)$ . Let  $\square$  denote a knowledge operator constructed from  $P_X$ .

We assume that Ann knows, or rather, we should say meta-knows, her model of knowledge. That is,  $P_X(x) = x$  where  $x$  represent the true underlying model of knowledge,  $(\Omega, P)$ . Let  $E_{\omega_1 \in P(\omega_2)}$  denote the event for which  $\omega_1 \in P(\omega_2)$ . For simplicity, we write  $\neg E_{\omega_1 \in P(\omega_2)}$  as  $E_{\omega_1 \notin P(\omega_2)}$

Assuming that  $(\Omega, P)$  is known to Ann, we infer that

$$x \in \square E_{\omega_1 \in P(\omega_2)} \text{ if and only if } \omega_1 \in P(\omega_2) \text{ in } (\Omega, P), \text{ represented by } x. \quad (1)$$

Meta-knowledge and knowledge, although modeled separately, are not independent entities. We are not able to express via  $(\Omega, P)$  all of the information included in  $(X, P_X)$ , but we can capture, in the underlying model of knowledge, some properties of  $K$  derived from the analysis of meta-knowledge. What helps is the relationship between models of knowledge and meta-knowledge as expressed in (1). Our conclusion is that the agent’s meta-knowing  $(\Omega, P)$  implies that in  $(\Omega, P)$  the agent knows whether the knowledge axioms are true.

### Proposition 4.1

*If the agent meta-knows her model of knowledge, then she knows whether the knowledge axioms hold.*

**Proof:**

We can formally state the proposition in the following way. If  $P_X(x) = x$ , where  $x$  is identified with

$(\Omega, P)$ , then (see Section 3) for each state  $\omega \in \Omega$  and each event,  $E \subset \Omega$ , (a)  $\omega \in K(\neg KE \cup E) \cup K(KE \cap \neg E)$ , (b)  $\omega \in K(\neg KE \cup KKE) \cup K(KE \cap \neg KKE)$ , and (c)  $\omega \in K(KE \cup K\neg KE) \cup K(\neg KE \cap \neg K\neg KE)$ .

First, we show that Ann knows whether the Truth Axiom holds. By assumption in (1), for each  $\omega$  we have either  $\Box E_{\omega \in P(\omega)}$  or  $\Box E_{\omega \notin P(\omega)}$ . That is, at the meta-level, Ann knows whether the Truth Axiom ( $\omega \in P(\omega)$ ) is true. Since we know how to translate that meta-result in the language of  $(\Omega, P)$ , we infer that, for all  $\omega$  and for all  $E$ , we have  $\omega \in K(\neg KE \cup E) \cup K(KE \cap \neg E)$ . To show that meta-knowledge of  $(\Omega, P)$  implies knowing whether the Positive Introspection Axiom holds, we fix two states,  $\omega_1$  and  $\omega_2$ , such that  $\omega_1 \in P(\omega_2)$ . If  $\omega_3 \in P(\omega_2)$  whenever  $\omega_3 \in P(\omega_1)$ , then the Positive Introspection Axiom holds. Since Ann meta-knows these relations, she meta-knows whether the Positive Introspection Axiom is satisfied. As we did with the analysis of the Truth Axiom, we translate our conclusion into  $(\Omega, P)$ : for all  $\omega$  and for all  $E$ , we have  $\omega \in K(\neg KE \cup KKE) \cup K(KE \cap \neg KKE)$ . Finally, we analyze the Negative Introspection Axiom. As previously, we assume that  $\omega_1 \in P(\omega_2)$ . If  $\omega_3 \in P(\omega_1)$  whenever  $\omega_3 \in P(\omega_2)$ , then the Negative Introspection Axiom holds. Thus, Ann meta-knows whether the axiom is true – a fact that we incorporate into  $(\Omega, P)$ . Thus, for all  $\omega$  and for all  $E$ , we have  $\omega \in K(KE \cup K\neg KE) \cup K(\neg KE \cap \neg K\neg KE)$ . ■

Combined Propositions 3.1 and 4.1 constitute the main result:

*Assuming that the agent meta-knows her model of knowledge and satisfies the Truth Axiom implies that she knows that she is not epistemically boundedly rational.*

That is, she knows – correctly – that all knowledge axioms hold. In other words, if the model is known to Ann, who is epistemically boundedly rational, then the model becomes inconsistent. Our result indicates that in the long-standing debate in economics about agents knowing or not knowing the model, it is safer to assume the latter.

Note also that the inverse claim – if knowledge axioms hold, then Ann knows about – is a well-known result in epistemics.

## 5 Applications

Our main result is that the agent must satisfy all knowledge axioms if she knows her own model of knowledge and the Truth Axiom holds. More generally, assuming the agents know the model changes the model. Next, we discuss two applications of our result.

## 5.1 Meta-knowledge and agreeing to disagree

The Agreement Theorem (Aumann [1]) shows that under assumption of agents having possibility correspondences satisfying all knowledge axioms and their priors being the same, their posteriors are also the same if they are commonly known. Does this result require agents' partitions to be commonly known (in the sense of meta-knowledge)? The answer is no. Since knowledge about possibility correspondence is incorporated into the meta-model, we know that agents knowing their own possibility correspondences must, by implication, satisfy the knowledge axioms. However, Samet [23] shows that the agents will not agree to disagree, even if their possibility correspondences violate the Negative Introspection Axiom. Our result indicates that the agents in Samet's paper do not know their own possibility correspondences. In consequence, possibility correspondences need not be either commonly known or known for the Agreement Theorem to hold.

## 5.2 Meta-knowledge and the no-trade theorems

If agents are characterized by risk-aversion and common priors, and their possibility correspondences satisfy knowledge axioms, then, as showed by Milgrom and Stokey [19], we obtain the no-trade theorem. One solution to this result was proposed by Morris [20], who advocates removal of the common prior assumption. Alternatively, Geanakoplos [11] shows that trade occurs if we assume violation of knowledge axioms. These two approaches are indeed equivalent, as shown by Brandenburger et al. [5]. However, if we require agents to know the model, then only non-common priors can help us avoid the no-trade theorem. Since knowledge is included in the setup, knowing the model implies that agents do not violate knowledge axioms, and, with common priors, we return to the Milgrom-Stokey construction.

## References

- [1] Aumann, R., "Agreeing to Disagree," *Annals of Statistics*, 4, 1976, 1236-1239.
- [2] Aumann, R., "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica*, 55, 1987, 1-18.
- [3] Aumann, R., "Interactive Epistemology I: Knowledge," *International Journal of Game Theory*, 28, 1999, 263-300.
- [4] Brandenburger, A., "Knowledge and Equilibrium in Games," *Journal of Economic Perspectives*, 6(4), 1992, 83-101.

- [5] Brandenburger, A., E. Dekel, and J. Geanakoplos, "Correlated Equilibrium with Generalized Information Structures," *Games and Economic Behavior*, 4, 1992, 182-201.
- [6] Brandenburger, A., and E. Dekel, "Hierarchies of Beliefs and Common Knowledge," *Journal of Economic Theory*, 59, 1993, 189-198.
- [7] Chellas, B.F., *Modal Logic. An Introduction.*, Cambridge University Press, 1980.
- [8] Dekel, E., and F. Gul, "Rationality and Knowledge in Game Theory," in Kreps, D.M., and K.F. Wallis (eds.), *Advances in Economics and Econometrics: Theory and Applications*, volume 1, Seventh World Congress, Cambridge University Press, 1997.
- [9] Fagin, R., J.Y. Halpern, Y. Moses, and M.Y. Vardi, *Reasoning about Knowledge*, MIT Press, 1995.
- [10] Fagin, R., J. Geanakoplos, J.Y. Halpern, and M.Y. Vardi, *The Hierarchical Approach to Modeling Knowledge and Common Knowledge*, *International Journal of Game Theory*, 28, 1999, 331-365.
- [11] Geanakoplos, J., "Game Theory Without Partitions, and Applications to Speculation and Consensus," Cowles Foundation Discussion Paper No. 914, Cowles Foundation for Research in Economics, Yale University, 1989.
- [12] Geanakoplos, J., "Common Knowledge," *Journal of Economic Perspectives*, 6(4), 1992, 53-82.
- [13] Gilboa, I., "Information and Meta Information," in M.Y. Vardi (ed.), *Proceedings of TARK II*, 1988, 227-243.
- [14] Harsanyi, J., "Games with Incomplete Information Played by 'Bayesian' Players, I-III," *Management Science*, 14, 1967/68, 159-182, 320-334, 486-502.
- [15] Hart, S., A. Heifetz, and D. Samet, "'Knowing Whether,' 'Knowing That,' and The Cardinality of State Space," *Journal of Economic Theory*, 70, 1996, 249-256.
- [16] Heifetz, A., "The Bayesian Formulation of Incomplete Information – The Non-Compact Case," *International Journal of Game Theory*, 21, 1993, 329-338.
- [17] Heifetz, A., "Non-well-founded-Type Spaces," *Games and Economic Behavior*, 16, 1996, 202-217.
- [18] Mertens, J.F., and S. Zamir, "Formulation of Bayesian Analysis for Games with Incomplete Information," *International Journal of Game Theory*, 14, 1985, 1-29.
- [19] Milgrom, P. and N. Stokey, "Information, Trade and Common Knowledge," *Journal of Economic Theory*, 26, 1982, 17-27.

- [20] Morris, S., "Trade with Heterogenous Prior Beliefs and Asymmetric Information," *Econometrica*, 62, 1994, 1327-1347.
- [21] Reny, P.J., "Rationality in Extensive-Form Games," *Journal of Economic Perspectives*, 6(4), 1992, 103-118.
- [22] Rubinstein, A., *Modeling Bounded Rationality*, MIT Press, 1998.
- [23] Samet, D., "Ignoring Ignorance and Agreeing to Disagree," *Journal of Economic Theory*, 52, 1990, 190-207.
- [24] Samet, D., "The Sure-thing Principle and Independence of Irrelevant Knowledge," working paper, Tel Aviv University, 2008.
- [25] Samuelson, L., "Modeling Knowledge in Economic Analysis," *Journal of Economic Literature*, 42(2), 2004, 367-403.