A PRIORI AND A POSTERIORI KNOWLEDGE IN EPISTEMIC LOGIC

Abstract. We study the dichotomy of a priori and a posteriori in the multi-agent epistemic logic S5. A formula φ is said to be *a posteriori discernable* by an individual i if in some possible world i does not know weather φ. A formula is said to be *a priori discernable* by i if in all possible worlds i knows whether φ. We show that the formulas that are a priori discernable by i are theorems, contradictions, and formulas that are logically equivalent to a description of i's knowledge. The formulas known by i in a given possible world is split into two parts: The formulas that i *a priori knows* — i.e., the a priori discernable formulas by i that i knows in this possible world, and the formulas that i *a posteriori knows* — i.e., the a posteriori discernable formulas by i that i knows in this possible world, We characterize these two types of knowledge and show that a posteriori knowledge can be retrieved from a priori knowledge and vice versa.

1. Introduction

The distinction between *a priori* and *a posteriori* knowledge goes back to antiquity and played an important role later in Immanuel Kant's Critique of Pure Reason. Roughly speaking, a priori knowledge is logically necessary and it is obtained by reasoning and deduction without any experience and observation. The knowledge of tautologies and more generally of mathematics is considered by most students of epistemology as a priori knowledge. In contrast, a posteriori knowledge concerns facts that are logically contingent, and can be acquired only by experiencing the world and observing it.

A priori and a posteriori discernability. We first note that the notions of a priori and a posteriori can be applied to statements rather than the *knowledge* of statements. A simple example demonstrates this. So far we do not know whether there is life on Mars. If we know in the future that there is life on Mars, or if we know that there is not, this knowledge will obviously be a posteriori knowledge. But even before we know one or the other, the pair of statements that there is life on Mars and that there is not, are of the type of statements that require observation and experiment to be known. We refer to this kind of statements as *a posteriori discernable*. Discernability, as opposed to knowledge, involves two statements, a statement and its negation, and it is about knowing *whether* rather than knowing *that*.[1]

Date: April 22, 2024.

[1]The operator "knowing whether" and its properties were studied in Hart, Heifetz and Samet (1996).

We study the notions of a priori and a posteriori in a propositional multi-agent epistemic logic. Formulas in this logic are generated by atomic formulas, logical connectives and for each individual i in a set I of individuals a knowledge operator K_i. Thus, if φ is a formula in the language, then $K_i\varphi$ is the formula that says that i knows φ. We assume that the knowledge operators satisfy the axioms of the propositional modal logic S5. In particular knowledge of each individual i satisfies for each formula φ the *truth axiom* $K_i\varphi \to \varphi$ and the two axioms $K_i\varphi \to K_iK_i\varphi$, and $\neg K_i\varphi \to K_i\neg K_i\varphi$. By the truth axiom we infer from these axioms $K_i\varphi \leftrightarrow K_iK_i\varphi$, and $\neg K_i\varphi \leftrightarrow K_i\neg K_i\varphi$.

The first difficulty we face in formalizing the dichotomy of a priori/a posteriori in such a simple logic is the lack of any notion of observability or experience in this logic. However, the semantics of the logic is built on the notion of possible worlds that can serve as a surrogate to the missing notion of observability. In the example above we noticed that we do not know whether there is life on Mars or not. This is so, because not enough observations have been made to answer the question. The lack of observation is demonstrated by the fact that there is a possible world, in fact the one we live in, in which we do not know whether or not there is life on Mars. Conversely, the existence of a possible world in which one cannot tell whether there is life on Mars shows that something is required for knowing that. We can call this missing something, an "observation" of the world.

This leads us to the following definition of a posteriori discernability.

> A formula φ is *a posteriori discernable* by an individual i if in some possible world i does not know whether φ. That is, $\neg K_i\varphi \wedge \neg K_i\neg\varphi$ holds true in some possible world.

By the completeness theorem for this logic, we can equivalently define a posteriori discernability in terms of provability. The formula φ is a posteriori discernable by individual i if the formula $\neg K_i\varphi \wedge \neg K_i\neg\varphi$, is consistent, that is, its negation that says that i knows whether φ, $K_i\varphi \vee K_i\neg\varphi$, is not a theorem.

A formula is a priori discernable if it is not a posteriori discernable, namely,

> A formula φ is *a priori discernable* by an individual i if in all possible worlds i knows whether φ. That is, $K_i\varphi \vee K_i\neg\varphi$ holds true in all possible worlds.

Again, a priori discernability can be defined equivalently by provability: the formula φ is a priori discernable if $K_i\varphi \vee K_i\neg\varphi$ is a theorem of the logic.

A priori and a posteriori knowledge. For each individual, the set of all formulas is partitioned into two parts: a priori and a posteriori discernable formulas. This partition is independent of the possible world. The actual knowledge of an individual depends of course on the possible world. Thus, for example, an individual may know φ in one possible world, $\neg\varphi$ in another possible world and know neither φ nor $\neg\varphi$ in yet another possible world.

We refer to the set of formulas that are known by an individual i in a given possible world as i's *ken* in this world.[2] We characterize sets of formulas that can be the ken of an individual in some possible world.

The partition of all formulas into a priori and a posteriori formulas discernable by i induces a partition of each ken of i into formulas that i knows and are a priori discernable by i and formulas that i knows and are a posteriori discernable by i. We can now define a priori and a posteriori *knowledge* in each possible world.

> For φ in the ken of i in a given possible world ω, we say that *i knows a priori* φ at ω if φ is a priori discernable by i, and say that *i knows a posteriori* φ at ω if φ is a posteriori discernable by i.

Examples.

(1) An atomic formula p is clearly a posteriori discernable by any individual i since there is a possible world in which $\neg K_i p \wedge \neg\, K_i \neg p$ holds true. That is, in this world i knows neither p nor $\neg p$.

(2) It can also easily be shown that $K_j p$, for individual $j \neq i$, is a posteriori discernable by i as there is a possible world in which $\neg K_i K_j p \wedge \neg K_i \neg K_j p$ holds. This reflects the fact that i cannot gain knowledge of $K_j p$ by contemplation alone. But observing j's observation of p is a good experimental reason for i to know $K_j p$.

(3) If φ is a theorem of the logic, then by the generalization inference rule in our logic $K_i\varphi$ is a theorem, and in particular $K_i\varphi \vee K_i\neg\varphi$ is a theorem, and hence φ is a priori discernable by i, as one may expect.

(4) The less trivial example of a priori discernability by i is the formula $K_i\varphi$. Obviously, in each possible world either $K_i\varphi$ holds or $\neg K_i\varphi$ holds. If $K_i\varphi$ holds true it follows from the axiom $K_i\varphi \to K_iK_i\varphi$ that $K_iK_i\varphi$ holds true. If $\neg K_i\varphi$ holds true it follows from the axiom $\neg K_i\varphi \to K_i\neg K_i\varphi$ that $K_i\neg K_i\varphi$ holds true. Thus, in each possible world $K_iK_i\varphi \vee K_i\neg K_i\varphi$ holds true, which renders $K_i\varphi$ and $\neg K_i\varphi$ a priori discernable by i. Hence, when i knows one of these formulas, his knowledge is a priori. We discuss this example next.

Introspection? Most epistemologists consider the formula $K_iK_i\varphi$ a posteriori knowledge of i. This is so because for i to know that she knows φ, i has to introspect her mental state. For example, Russell (2020) in his summary of the state-of-the-art of the study of a priori describes the observations that are needed to justify a posteriori knowledge:

> Observations based on our senses, or introspection about our current mental state, are needed for us to be empirically, or *a posteriori*, justified in believing that some proposition is true.

[2]The term was first used in this sense in Samet (1990).

Thus, according to this view, one's mental state is like the state of the outer world, except that to examine it one has to look inside, that is, introspect. In this spirit, some call the axioms $K_i\varphi \to K_iK_i\varphi$ and $\neg K_i\varphi \to K_i\neg K_i\varphi$ positive and negative *introspection*, respectively.

But here we are studying the logic S5 and in this logic knowledge does not seem to be acquired by introspection or observation of some mental state. If such introspection was required, then a state of affairs where one cannot tell which of $K_i\varphi$ and $\neg K_i\varphi$ is the case would be possible, at least in the split second before the introspection was made. However, such a state of affairs is impossible in S5 because $\neg K_iK_i\varphi \wedge \neg K_i\neg K_i\varphi$ is a contradiction in S5. One does not reach the conclusion that one knows or that one does not know φ as a result of time-consuming soul searching. Rather, these axioms in our logic being theorems, or equivalently being true in all possible worlds, *define* the meaning of knowledge.[3] Hintikka (1962) made a similar argument in explaining the positive introspection axiom. The same argument can be made for the negative introspection axiom.

Characterizing a priori. Since the set of a posteriori discernable formulas is the complement of the set of a priori discernable formulas, the characterization of the latter set characterizes the former set.

If either φ or $\neg\varphi$ is a theorem, then by the generalization rule either $K_i\varphi$ or $K_i\neg\varphi$ is a theorem and thus $K_i\varphi \vee K_i\neg\varphi$ is a theorem. Therefore all theorems and contradictions are a priori discernable by i. Example (4) shows that there are other formulas that are a prior discernable. We generalize this example. We say that a formula *describes i's knowledge* if it is generated by logical connectives from formulas of the form $K_i\varphi$.

We characterize the formulas that are a priori discernable by i as follows:

> A formula is a priori discernable by i if and only if it is either a theorem or a contradiction, or it is logically equivalent to a formula that describes i's knowledge.

Each ken of i is partitioned into the set of formulas that i knows a priori and the set of formulas that i knows a posteriori. We characterize the two parts of kens. This characterization results in full determinacy of each part by the other part.

> The a priori knowledge part of a ken of i determines its a posteriori knowledge part, and vice versa, the a posteriori knowledge part of a ken of i determines its a priori knowledge part.

Discussion. Possible worlds play an important role in this paper. First, they are used to define a priori and a posteriori discernability, and second,

[3]Russell (2020) considers the sentence "If you know something, you believe it and it's true" a priori knowledge. One does not need to examine her mind to see that she believes when she knows. It follows from the definitions of belief and knowledge. Similarly the axioms of knowledge in S5 define knowledge rather than describing a psychological process.

they define the kens of an individual, which vary with possible worlds. In the context of the a priori, possible worlds were first introduced in Kripke (1980, pp. 35-39). Kripke distinguishes between a priori and necessary. The first, he claims, is an epistemic notion, and the second, a metaphysical one that has nothing to do with knowledge. Nevertheless, he admits that the two notions are close and perhaps even identical, but says that such a claim requires a proof. Kripke sketches the relation between the a priori and the necessary but raises doubts as to how well this explains the relation.

> ... if something not only happens to be true in the actual world but is also true in all possible worlds, then, of course, just by running through all the possible worlds in our heads, we ought to be able with enough effort to see, if a statement is necessary, that it is necessary, and thus know it a priori. But really this is not so obviously feasible at all (Kripke, 1980, p. 38).

Possible worlds in this paragraph play a role in defining the modalities of possibility and necessity. Epistemology is manifested in Kripke's analysis by thinking about the non-epistemic possible worlds but it is not a modality.

Possible worlds also paly an important role in our analysis. However, unlike Kripke, we consider knowledge as a modality. The semantics of the modality 'knowledge' is defined in terms of possible worlds, but the logic is purely epistemic and the modality of necessity plays no role. In the epistemic logic we distinguish between a priori and a posteriori discernability, which is a property of statements, and a priori a posteriori knowledge, which varies with possible worlds and is not fixed like in Kripke's analysis.

Kripke's concern about the feasibility of thinking of all possible worlds is justified as there is a continuum of possible worlds.[4] However, this concern is lessened by the fact that a formula is true in *all* possible worlds if and only if it is true in each of the finitely many possible worlds in some specific model.[5] Indeed, this fact is proved here where the finite model is the n-canonical model with n being the depth of the formula.

The logic of the knowledge operators studied here, known as the modal logic S5, raised some objection by Hintikka (1962), who considered the combination of the axioms of truth and negative introspection to be incoherent. However, Stalnaker (2006) who agreed with Hintikka's criticism, observed that students of interaction, like game theorists and computer scientists, use S5 almost exclusively. The reason for studying this logic here is its prevalence and simplicity. The results obtained here are not trivially extended to other logics of knowledge and belief. Other logics may require a modification of the definitions of a posteriori and a priori knowledge, and the properties of these types of knowledge will surely vary with the different logics.

[4]See, Hart, Heifetz and Samet (1996) and Aumann (1999).

[5]This is Theorem 3.2.2 in Fagin, Halpern, Moses and Vardi (1995).

2. Preliminaries

2.1. **Syntax.** We consider a logic of multi-agent knowledge for a finite set I of individuals.[6] Starting with a set A of *atomic formulas*, the set of *formulas*, $\mathcal{F}$ is the smallest set that contains A and for each φ and ψ in $\mathcal{F}$, contains $\neg\varphi$ (read, not φ), $(\varphi \to \psi)$ (read, if φ then ψ), and $K_i\varphi$ (read, i knows φ). The *logical connectives* $\wedge$, $\vee$ and $\leftrightarrow$ are defined as usual in terms of $\neg$ and $\to$.

The set of theorems in $\mathcal{F}$ is defined inductively starting with a set of sentences called *axioms*. The set of axioms consists of all propositional calculus tautologies and for each i and sentences φ and ψ, each of the following sentences:

(K) $K_i(\varphi \to \psi) \to (K_i\varphi \to K_i\psi)$;
(T) $K_i\varphi \to \varphi$ (truth axiom);
(5) $\neg K_i\varphi \to K_i\neg K_i\varphi$ (negative introspection).

The set of *theorems* $\mathcal{T}$ is the smallest set of formulas that contains all the axioms, and for each $\varphi \in \mathcal{T}$, if $\varphi \to \psi \in \mathcal{T}$ then $\psi \in \mathcal{T}$ (*modus ponens*), and for each i, $K_i\varphi \in \mathcal{T}$ (*generalization*). For a theorem φ we write $\vdash \varphi$. Two useful theorems are $K_i\varphi \to K_iK_i\varphi$ (positive introspection) and $K_i(\varphi\wedge\psi) \leftrightarrow (K_i\varphi \wedge K_i\psi)$. A negation of a theorem is called a *contradiction*. A sentence that is not a contradiction is *consistent*. A formula is *inconsistent* when it is not consistent, that is, it is a contradiction. When $\vdash \varphi \leftrightarrow \psi$ we say that φ and ψ are *logically equivalent*, or *equivalent* for short, and write $\varphi \equiv \psi$. If we replace a subformula of φ with an equivalent formula, the resulting formula is equivalent to φ. A set of formulas Φ is *consistent* if there are no formulas $\varphi_1, \ldots, \varphi_n$ in Φ such that $(\varphi_1 \wedge \cdots \wedge \varphi_n)$ is inconsistent.

2.2. **Models.** We use *Kripke models*, *models* for short, as the semantics of the logic. A *model* consists of a set Ω of *possible worlds*; an equivalence relation on Ω, $\sim_i$ for each individual i; and *interpretation* of each atomic formula p, which is a subset of Ω denoted by $[p]$. We denote by Π_i the partition of Ω into equivalence classes of $\sim_i$, and by $\Pi_i(\omega)$ the equivalence class containing ω.

The interpretation of all formulas is defined by induction on the structure of formulas; $[\neg\varphi] = \Omega\setminus[\varphi]$; $[\varphi\wedge\psi] = [\varphi]\cap[\psi]$; and $[K_i\varphi] = \{\omega \mid \Pi_i(\omega) \subseteq [\varphi]\}$. We say that φ *is true in* ω if $\omega \in [\varphi]$. It is easy to show that the logic is *sound* for the family of Kripke models, that is, each theorem is true in each possible world of each model. In other words, if φ is a theorem then in each model Ω, $[\varphi] = \Omega$.

2.3. **The canonical model.** Lindenbaum's Lemma claims that if Φ is a consistent set of formulas such that $\varphi \notin \Phi$ and $\neg\varphi \notin \Phi$, then at least one of $\Phi\cup\{\varphi\}$ and $\Phi\cup\{\neg\varphi\}$ is consistent. A *maximal consistent set* is a consistent set of formulas Φ such that each proper superset of Φ is inconsistent. By

[6] This logic is described in detail in Fagin, Halpern, Moses and Vardi (1995)

Lindenbaum's Lemma, Φ is maximal consistent if and only if it is consistent and for each formula φ, exactly one of φ and $\neg\varphi$ is in Φ. Moreover, by this lemma, every consistent set Ψ is a subset of some maximal consistent set. Since a formula φ is a theorem if and only if $\neg\varphi$ is inconsistent, it follows that φ is a theorem if and only if it is contained in each maximal consistent set.

Let Ω^∞ be the set of all maximal consistent sets. We make Ω^∞ a model, which we call the *canonical model* by defining for each atomic formula p, $[p] = \{\omega \mid p \in \omega\}$, and for each ω and ω' in Ω^∞ and i, $\omega \sim_i \omega'$ whenever for each formula φ, $K_i\varphi \in \omega$ if and only if $K_i\varphi \in \omega'$. The special feature of the canonical model is the following property that is proved by induction of the structure of formulas: For each formula φ,

$$[\varphi] = \{\omega \in \Omega^\infty \mid \varphi \in \omega\}. \tag{1}$$

Note that if φ is true in every possible world in each model, then in particular, for the canonical model $[\varphi] = \Omega^\infty$. Thus, by equation (1), $\varphi \in \omega$ for each $\omega \in \Omega^\infty$, which implies, as we argued before, that φ is a theorem. This proves that the logic is *complete* for the canonical model and a fortiori for the family of Kripke models.[7]

2.4. ***n*-canonical models.** The *depth of a formula* φ, denoted $\mathrm{dep}(\varphi)$, measures the number of nested knowledge operators in φ. Formally, dep is defied by induction of the structure of formulas. For each atomic formula p, $\mathrm{dep}(p) = 0$, $\mathrm{dep}(\neg\varphi) = \mathrm{dep}(\varphi)$, $\mathrm{dep}(\varphi \wedge \psi) = \max(\mathrm{dep}(\varphi), \mathrm{dep}(\psi))$, and for each individual i, $\mathrm{dep}(K_i\varphi) = \mathrm{dep}(\varphi) + 1$.

An *n-maximal consistent set* is a consistent set of formulas of depth $\leq n$, Φ, such that each proper superset of Φ that consists of formulas of depth $\leq n$ is inconsistent. Equivalently, a set Φ of formulas of depth $\leq n$ is n-maximal consistent if it is consistent and for each φ with $\mathrm{dep}(p) \leq n$ exactly one of φ and $\neg\varphi$ is in Φ. The set of all n-maximal consistent sets is denoted by Ω^n. We make Ω^n a model, which we call the *n-canonical model*, by defining $[p]$ and $\sim_i$ exactly as they are defined for Ω^∞. Similarly to (1) and with the same proof we can show that for each formula φ with $\mathrm{dep}(\varphi) \leq n$,

$$[\varphi] = \{\omega \in \Omega^n \mid \varphi \in \omega\}. \tag{2}$$

The relation between Ω^∞ and Ω^n is simple.

Claim 1. *For every $n \geq 1$, each $\omega \in \Omega^n$ is contained in some $\omega' \in \Omega^\infty$ (indeed in many such worlds), and each $\omega' \in \Omega^\infty$ contains exactly one world $\omega \in \Omega^n$.*

The first part follows from Lindenbaum's Lemma and the second from ω' being a maximal consistent set. Since a formula φ is a theorem if and only if $\varphi \in \omega$ for each $\omega \in \Omega^\infty$ it follows by Claim 1 that

[7] This is how the completeness theorem for this logic is proved in Fagin, Halpern, Moses and Vardi (1995) (see theorem Theorem 3.1.3 there).

Claim 2. *A formula φ with* $\mathrm{dep}(\varphi) \leq n$ *is a theorem if and only if $\varphi \in \omega$ for each $\omega \in \Omega^n$.*

Finally, in the n-canonical model, for each i, each equivalence class of $\sim_i$ can be expressed by a formula of degree $\leq n$ that describes i's knowledge in the possible worlds of the equivalence class.

Claim 3. *For $\omega \in \Omega^n$ with $n \geq 1$ and individual i, let*

$$\varphi^i_\omega = (\wedge_{K_i\varphi\in\omega} K_i\varphi) \bigwedge (\wedge_{\neg K_i\varphi\in\omega} \neg K_i\varphi).$$

Then, $\Pi_i(\omega) = [\varphi^i_\omega]$.

Indeed, by the definition of $\sim_i$, for any $\omega' \sim_i \omega$, $\varphi^i_\omega \in \omega'$, and therefore, by (2), $\Pi_i(\omega) \subseteq [\varphi^i_\omega]$. If $\omega' \not\sim_i \omega$, then $\varphi^i_\omega \neq \varphi^i_{\omega'}$ and moreover, they are contradictory. Hence, $\varphi^i_\omega \notin \omega'$, which establishes the equality of the claim.

The n-canonical model is very similar to the well-known canonical model. Nevertheless, we cannot give a reference to this notion in the literature of modal logic and knowledge. There are several works that define set theoretically a sequence of finite models for modal logic by taking in each step certain supersets of the possible worlds of the previous step. The possible worlds in this construction are called in Fagin and Vardi (1985) and Fagin, Halpern and Vardi (1991) *k-ary worlds* and the limit of such worlds is called there a *modal structure*. A similar construction is carried out in Heifetz and Samet (1999). It is possible to show that the model of k-ary worlds is isomorphic to the k-canonical model and the model of modal structures is the canonical model. However, these works do not make this connection.

3. A priori and a posteriori discernability

For each individual i, we divide the set of all formulas into two subsets: the set of formulas that are *a posteriori discernable by individual* i, and the complement, the set of formulas that are *a priori discernable by individual* i. Intuitively, a formula φ is a posteriori discernable by individual i if telling which of φ and $\neg\varphi$ is true requires an observation. Although observation cannot be expressed directly in our simple language, we can think of a possible world in which the individual cannot tell which of φ and $\neg\varphi$ is true as a world in which no observation is made. A formula is a priori discernable by individual i if it is not a posteriori discernable by i. That is, if in all the possible worlds i knows whether the formula or its negation holds. Note that both a priori and a posteriori discernability are defined in terms of knowing *whether*, rather than knowing *that*. In particular, each of the sets of a posteriori and a priori discernable formulas is closed under negation. Formally,

Definition 1. *A formula φ is* a posteriori discernable by individual i *if in some possible world in some model the formula $(\neg K_i\varphi) \wedge (\neg K_i\neg\varphi)$ holds true. Equivalently, φ is a posteriori discernable by individual i if $(\neg K_i\varphi) \wedge (\neg K_i\neg\varphi)$ is consistent.*

A formula φ *is* a priori discernable by individual i *if in all possible worlds of each model the formula* $(K_i\varphi) \vee (K_i\neg\varphi)$ *holds true . Equivalently,* φ *is a priori discernable by individual i if* $(K_i\varphi) \vee (K_i\neg\varphi)$ *is a theorem.*

We denote by $\mathrm{R_i}$ the set of formulas that are a priori discernable by i and by $\mathrm{S_i}$ the set of formulas that are a posteriori discernable by i. Thus,

$$\mathrm{R_i} = \{\varphi \mid \vdash (K\varphi) \vee (K\neg\varphi)\} \qquad \mathrm{S_i} = \{\varphi \mid\not\vdash (K\varphi) \vee (K\neg\varphi)\}.$$

Obviously, for each atomic formula p, p and $\neg p$ are in $\mathrm{S_i}$ as there is a simple model and a possible world in the model in which i knows neither p nor $\neg p$. If φ is a theorem then $K_i\varphi$ is a theorem and therefore $K_i\varphi \vee K_i\neg\varphi$ is a theorem, and thus, φ and $\neg\varphi$ are in $\mathrm{R_i}$. For each formula φ, $K_i\varphi$ and $\neg K_i\varphi$ are in $\mathrm{R_i}$ since $(K_iK_i\varphi) \vee (K_i\neg K_i\varphi)$ is a theorem.

The next proposition provides us with alternative definitions of $\mathrm{R_i}$ in terms of provability.

Proposition 1. *The following conditions are equivalent.*

(1) $\vdash K_i\varphi \vee K_i\neg\varphi$.
(2) $\vdash K_i\neg\varphi \leftrightarrow \neg K_i\varphi$.
(3) $\vdash \varphi \rightarrow K_i\varphi$;
(4) $\vdash \varphi \leftrightarrow K_i\varphi$;
(5) $\vdash \neg\varphi \rightarrow K_i\neg\varphi$;
(6) $\vdash \neg\varphi \leftrightarrow K_i\neg\varphi$;

Proof: (1) and (2) are equivalent: By (2) $\vdash \neg K_i\varphi \rightarrow K_i\neg\varphi$ which is (1). The other implication in (2) is obtained by the truth axiom as $\vdash K_i\neg\varphi \rightarrow \neg\varphi$, and $\vdash \neg\varphi \rightarrow \neg K_i\varphi$. **(3) and (4) are equivalent and (5) and (6) are equivalent:** Obviously, (4) implies (3) and (6) implies (5). The other implications in (4) and (6) are just the truth axiom. **(1) and (3) are equivalent:** By the truth axiom $\vdash \varphi \rightarrow \neg K_i\neg\varphi$. Together with (1) we conclude (3) by propositional calculus. For the converse, (3) implies $\vdash \neg K_i\varphi \rightarrow \neg\varphi$ and hence by generalization $\vdash K_i(\neg K_i\varphi \rightarrow \neg\varphi)$. By axiom K and negative introspection $\vdash \neg K_i\varphi \rightarrow K_i\neg\varphi$, which is (1). **(1) and (5) are equivalent:** The proof is similar to the equivalence of (1) and (3). ∎

In the following theorem we show that $\mathrm{R_i}$ consists of two types of formulas, trivial ones and formulas that describe i's knowledge.

Definition 2. *A formula* φ *is said to be* trivial *if either* $\vdash \varphi$ *or* $\vdash \neg\varphi$.

To describe i's knowledge we consider $\mathcal{F}_i$, the set of formulas generated by the logical connectives from formulas of the form $K_i\varphi$.

Theorem 1. *A formula* φ *is in* $\mathrm{R_i}$ *if and only if either* φ *is trivial or* $\varphi \equiv \psi$ *for some* $\psi \in \mathcal{F}_i$ *with* $\mathrm{dep}(\psi) \leq \mathrm{dep}(\varphi)$. [8]

[8] The 'or' in this theorem is not exclusive. For example $K_ip \vee \neg K_ip$ is in $\mathrm{R_i}$ and $\mathcal{F}_i$. It is both trivial and it is equivalent to itself. There are trivial formulas that do not satisfy the other condition, for example $p \vee \neg p$ for an atomic formula p, which is not equivalent

Proof: Let $\varphi \in \mathrm{R_i}$. We need to show that if φ is not trivial then there exists a formula ψ in $\mathcal{F}_i$ with $\mathrm{dep}(\psi) \leq \mathrm{dep}(\varphi)$ such that $\varphi \equiv \psi$. We show that if φ is not trivial and there is no such formula ψ, then $\varphi \notin \mathrm{R_i}$, that is $\varphi \in \mathrm{S_i}$.

Consider first the simple case where $\mathrm{dep}(\varphi) = 0$. In this case, just the fact that φ is not trivial implies that $\varphi \in \mathrm{S_i}$. Indeed, we can easily construct a model and a possible world in it at which $\neg K_i\varphi \wedge \neg K_i\neg\varphi$ holds true, that is, $\varphi \in \mathrm{S_i}$. Suppose, then, that $\mathrm{dep}(\varphi) = n > 0$. We show that $[\varphi]$ is not a union of equivalence classes of $\sim_i$ in the n-canonical model Ω^n. Indeed, for each equivalence class of $\sim_i$, π_i, chose $\omega \in \pi_i$. Then, by Claim 3, $\pi_i = [\varphi_\omega]$. Observe that $\varphi_\omega \in \mathcal{F}_i$ and $\mathrm{dep}(\varphi_\omega) \leq n$. If $[\varphi] = \cup_{k=1}^n \pi_i^k$, where each π_i^k is an equivalence class of $\sim_i$, then $[\varphi] = [\vee_{k=1}^n \varphi_{\omega^k}]$, where for each k, $\omega^k \in \pi_i^k$. Thus, $\varphi \leftrightarrow \vee_{k=1}^n \varphi_{\omega^k}$ is true in all possible worlds in Ω^n. Since the disjunction is of depth $\leq n$, it follows, by Claim 2, that $\varphi \equiv \vee_{k=1}^n \varphi_{\omega^k}$, which contradicts our assumption as $\vee_{k=1}^n \varphi_{\omega^k} \in \mathcal{F}_i$.

Since φ is not trivial, we conclude, by Claim 2, that $[\varphi]$ and $[\neg\varphi] = \Omega^n \setminus [\varphi]$ are not empty. Thus, since $[\varphi]$ is not a union of equivalence classes of $\sim_i$ there is an equivalence class of $\sim_i$ with non-empty intersection with $[\varphi]$ and with $[\neg\varphi]$. In each possible world in this equivalence class $\neg K_i\varphi \wedge \neg K_i\neg\varphi$ holds true, and thus $\varphi \in \mathrm{S_i}$.

To show the converse, we first note that if φ is trivial then by the generalization rule, $\vdash K_i\varphi \vee K_i\neg\varphi$, that is, $\varphi \in \mathrm{R_i}$. Next we claim that $\mathcal{F}_i \subseteq \mathrm{R_i}$. For this we use the following lemma.

Lemma 1. *For each $\varphi \in \mathcal{F}_i$, $\varphi \equiv K_i\varphi$.*

Proof: The proof is by induction on the structure of formulas in $\mathcal{F}_i$ The claim is obviously true for formulas $\varphi = K_i\psi$ by the truth axiom and positive introspection. Suppose $\varphi = \neg\psi$ and $\psi \equiv K_i\psi$. We need to show that $\neg\psi \equiv K_i\neg\psi$. Indeed, per our assumption and negative introspection, $\neg\psi \equiv \neg K_i\psi \equiv K_i\neg K_i\psi \equiv K_i\neg\psi$. Finally, suppose that $\varphi = \psi \wedge \xi$ and $K_i\psi \equiv \psi$ and $K_i\xi \equiv \xi$, then $K_i(\psi \wedge \xi) \equiv (K_i\psi) \wedge (K_i\xi) \equiv \psi \wedge \xi$. ∎

It follows from Lemma 1 and part (4) of Proposition 1 that $\mathcal{F}_i \subseteq \mathrm{R_i}$. To complete the proof of the theorem we observe that $\mathrm{R_i}$ is closed with respect to equivalence, and therefore if $\varphi \equiv \psi$ for some $\psi \in \mathcal{F}_i$, then $\varphi \in \mathrm{R_i}$. ∎

The proof of Theorem 1 delivers more than is stated in the theorem. We showed that the condition $\varphi \equiv \psi$ for some $\psi \in \mathcal{F}_i$, without the requirement $\mathrm{dep}(\psi) \leq \mathrm{dep}(\varphi)$, is enough to guarantee that $\varphi \in \mathrm{R_i}$. This requirement of the depth of ψ that seems to be a technical point is used in proving the non-technical result that a priori knowledge can be derived from a posteriori knowledge in part (b) of Theorem 2.

---

to any formula in $\mathcal{F}_i$ with lower or equal depth as there are no 0-depth formulas in $\mathcal{F}_i$. Obviously, there are formulas, like K_ip, which satisfy the other condition but they are not trivial.

4. A priori and a posteriori knowledge

A formula φ which is a posteriori discernable by i is not necessarily known by i in given possible world. In some possible worlds i may know φ, in some others i may know $\neg\varphi$ and in yet others i may know neither. The a priori knowledge of i may also depend on the possible world. Although $K_i p$ is a priori discernable by i, she may know $K_i p$ in some possible worlds and know $\neg K_i p$ in others. We now characterize the sets of formulas that i can know in a possible world and then characterize the intersection of such sets with $\mathrm{R_i}$ and $\mathrm{S_i}$.

The set of formulas that hold true in a possible world in some model is a maximal consistent set. Conversely, any maximal consistent set is a possible world in the canonical model. Thus, in order to characterize sets of formulas that are the formulas known by an individual in some possible world we can restrict ourselves to possible worlds in the canonical model. The set of formulas known by i in $\omega \in \Omega^\infty$ is called the *ken* of i in ω and is denoted by $\mathrm{ken}_i(\omega)$. That is, $\mathrm{ken}_i(\omega) = \{\varphi \mid K_i\varphi \in \omega\}$. We first characterize sets of formulas that are kens.

Proposition 2. *A set of formulas Φ is a ken of i if and only if:*

1. *Φ is consistent;*
2. *If $\varphi \in \Phi$ then $K_i\varphi \in \Phi$;*
3. *If $\varphi \notin \Phi$ then $\neg K_i\varphi \in \Phi$.*

Proof: Suppose $\Phi = \mathrm{ken}_i(\omega)$ for some $\omega \in \Omega^\infty$. (1) Since ω is consistent and $\mathrm{ken}_i(\omega) \subset \omega$, $\mathrm{ken}_i(\omega)$ is consistent. (2) If $\varphi \in \mathrm{ken}_i(\omega)$ then $K_i\varphi \in \omega$ and thus $K_iK_i\varphi \in \omega$, which implies $K_i\varphi \in \mathrm{ken}_i(\omega)$. (3) If $\varphi \notin \Phi$ then $K_i\varphi \notin \omega$ and therefore, $\neg K_i\varphi \in \omega$, which implies that $K_i\neg K_i\varphi \in \omega$ and thus $\neg K_i\varphi \in \mathrm{ken}_i(\omega)$.

Conversely, suppose that Φ satisfies (1)-(3). As Φ is consistent it can be extended to a maximal consistent set ω in Ω^∞. We show that $\Phi = \mathrm{ken}_i(\omega)$. If $\varphi \in \Phi$ then, by (2), $K_i\varphi \in \Phi \subseteq \omega$ and therefore, $\varphi \in \mathrm{ken}_i(\omega)$. Thus $\Phi \subseteq \mathrm{ken}_i(\omega)$. If $\varphi \in \mathrm{ken}_i(\omega)$ then $K_i\varphi \in \omega$. Suppose $\varphi \notin \Phi$, then by (3), $\neg K_i\varphi \in \Phi \subseteq \omega$, which contradicts the consistency of ω. Thus, $\mathrm{ken}_i(\omega) \subseteq \Phi$.∎

The partition of the set of formulas into a priori and a posteriori discernable formulas by i induces a partition of each of i's kens, $\mathbb{K}_i$ into the a priori and a posteriori known formulas: $\mathbb{K}_i^{\mathrm{R}} = \mathbb{K}_i \cap \mathrm{R_i}$, and $\mathbb{K}_i^{\mathrm{S}} = \mathbb{K}_i \cap \mathrm{S_i}$

In the next proposition we characterize these two sets.

Proposition 3. *For a set of formulas Φ,*

(a) *$\Phi = \mathbb{K}_i^{\mathrm{R}}$ for some ken of i, $\mathbb{K}_i$, if and only if Φ is a maximal consistent subset of $\mathrm{R_i}$;*
(b) *$\Phi = \mathbb{K}_i^{\mathrm{S}}$ for some ken of i, $\mathbb{K}_i$, if and only if $K_i\Phi \cup \neg K_i(S_i \setminus \Phi)$ is consistent.*

Proof: (a) Let $\mathbb{K}_i = \mathrm{ken}_i(\omega)$. As a subset of a consistent set, $\mathbb{K}_i^{\mathrm{R}}$ is consistent. Since ω is a maximal consistent set, for each $\varphi \in \mathrm{R_i}$ either $K_i\varphi \in \omega$

or $\neg K_i\varphi \in \omega$. In the first case, $\varphi \in \mathbb{K}_i^{\mathrm{R}}$. In the second case, as $\varphi \in R_i$ it follows, by part (2) of Proposition 1, that $K_i\neg\varphi \in \omega$, and hence $\neg\varphi \in \mathbb{K}_i^{\mathrm{R}}$. Thus, $\mathbb{K}_i^{\mathrm{R}}$ is a maximal consistent subset of $\mathrm{R_i}$.

Let Φ be a maximal consistent subset of $\mathrm{R_i}$. By part (4) of Proposition 1, for each $\varphi \in \Phi$, $K_i\varphi \equiv \varphi$ and therefore $K_i\varphi$ must be in Φ. Thus, $K_i\Phi \subseteq \Phi$. Since Φ is consistent it follows that $K_i\Phi$ is consistent. Hence there exists $\omega \in \Omega^\infty$ such that $K_i\Phi \subseteq \omega$. Let $\mathbb{K}_i = \mathrm{ken}_i(\omega)$. Then $\Phi \subseteq \mathbb{K}_i^{\mathrm{R}}$. Since Φ is a maximal consistent subset of $\mathrm{R_i}$ and, as we have shown, so is $\mathbb{K}_i^{\mathrm{R}}$, it follows that $\Phi = \mathbb{K}_i^{\mathrm{R}}$.

(b) Let $\Phi = \mathbb{K}_i^{\mathrm{S}}$ for a ken $\mathbb{K}_i$ of i. Then $\Phi \subseteq \mathbb{K}_i$ and therefore $K_i\Phi \subseteq \mathbb{K}_i$. Since $S_i \setminus \Phi$ is disjoint from $\mathbb{K}_i$ it follows by Propositions 2 that $\neg K_i(S_i \setminus \Phi) \subseteq \mathbb{K}_i$. As $\mathbb{K}_i$ is consistent, it follows that $K_i\Phi \cup \neg K_i(S_i \setminus \Phi)$ is consistent.

Conversely, suppose that $\Phi \subseteq S_i$ and $K_i\Phi \cup \neg K_i(S_i \setminus \Phi)$ is consistent, then $K_i\Phi \cup \neg K_i(S_i \setminus \Phi) \subseteq \omega$ for some $\omega \in \Omega^\infty$. Let $\mathbb{K}_i = \mathrm{ken}_i(\omega)$. Then $\Phi \subseteq \mathbb{K}_i \cap \mathrm{S_i} = \mathbb{K}_i^{\mathrm{S}}$. Also, $S_i \setminus \Phi$ is disjoint from $\mathbb{K}_i$ and therefore also from $\mathbb{K}_i^{\mathrm{S}}$. Hence, $\Phi = \mathbb{K}_i^{\mathrm{S}}$. ∎

The next theorem claims that for each ken of i, $\mathbb{K}_i$, each of the sets $\mathbb{K}_i^{\mathrm{R}}$ and $\mathbb{K}_i^{\mathrm{S}}$ determines the ken. That is, the a posteriori knowledge of i in a possible world can be derived from i's a priori knowledge at the possible world, and vice versa.

Theorem 2. *Let $\mathbb{K}_i$ and $\hat{\mathbb{K}}_i$ be kens of i.*

(a) *if $\mathbb{K}_i^{\mathrm{R}} = \hat{\mathbb{K}}_i^{\mathrm{R}}$, then $\mathbb{K}_i^{\mathrm{S}} = \hat{\mathbb{K}}_i^{\mathrm{S}}$;*

(b) *if $\mathbb{K}_i^{\mathrm{S}} = \hat{\mathbb{K}}_i^{\mathrm{S}}$, then $\mathbb{K}_i^{\mathrm{R}} = \hat{\mathbb{K}}_i^{\mathrm{R}}$.*

Proof:

(a) Consider $\varphi \in \mathrm{S_i}$. Since $K_i\varphi \in R_i$, it follows by Proposition 3 that one and only one of $K_i\varphi$ and $\neg K_i\varphi$ is in $\mathbb{K}_i^{\mathrm{R}}$. If $K_i\varphi \in \mathbb{K}_i^{\mathrm{R}}$ then $\varphi \in \mathbb{K}_i$ and thus, since $\varphi \in \mathrm{S_i}$, it follows that $\varphi \in \mathbb{K}_i^{\mathrm{S}}$. As $\mathbb{K}_i^{\mathrm{R}} = \hat{\mathbb{K}}_i^{\mathrm{R}}$, also $\varphi \in \hat{\mathbb{K}}_i^{\mathrm{S}}$. If $\neg K_i\varphi \in \mathbb{K}_i^{\mathrm{R}}$ then $\varphi \notin \mathbb{K}_i$ and thus $\varphi \notin \mathbb{K}_i^{\mathrm{S}}$, and also $\varphi \notin \hat{\mathbb{K}}_i^{\mathrm{S}}$. Thus, $\mathbb{K}_i^{\mathrm{S}} = \hat{\mathbb{K}}_i^{\mathrm{S}}$.

(b) In the next proposition we show that formulas in $\mathrm{R_i}$ can be expressed in terms of formulas in $\mathrm{S_i}$. Denote by $\mathcal{F}_i^S$ the subset of $\mathcal{F}_i$ that consists of the formulas generated by the logical connectives from the formulas of the form $K_i\varphi$ where $\varphi \in \mathrm{S_i}$.

Proposition 4. *If $\varphi \in \mathrm{R_i}$ and φ is not trivial then $\varphi \equiv \psi$ for some $\psi \in \mathcal{F}_i^S$.*

Proof: We first show the following simple claim.

Lemma 2. *If $\varphi \in \mathcal{F}_i$ is non-trivial, then $\varphi \equiv \varphi'$ for $\varphi' \in \mathcal{F}_i$ that is generated by logical connectives from formulas of the form $K_i\psi$ where ψ is non-trivial.*

Proof: The proof is by induction on the structure of formulas in $\mathcal{F}_i$. Obviously, if $\varphi = K_i\psi$ or $\varphi = \neg K_i\psi$ is non-trivial, then ψ is non-trivial and $\varphi' = \varphi$. Suppose $\varphi = K_i\psi_1 \wedge K_i\psi_2$ is non-trivial. None of ψ_1 and ψ_2 can be a contradiction because then φ is a contradiction. It is impossible that both ψ_1 and ψ_2 are theorems because then φ is a theorem. Thus, either ψ_1 and ψ_2 are both non-trivial, and $\varphi' = \varphi$, or one, say ψ_1, is a theorem and ψ_2 is non-trivial, in which case $\varphi' = K_i\psi_2$. ∎

By Theorem 1, if $\varphi \in \mathrm{R_i}$ and φ is non-trivial, then $\varphi \equiv \psi$ for some $\psi \in \mathcal{F}_i$, and obviously, as φ is not trivial ψ is not trivial. Thus, it is enough to show that any non-trivial formula in $\mathcal{F}_i$ is equivalent to a formula in $\mathcal{F}_i^S$. We prove it by induction on the depth of non-trivial formulas in $\mathcal{F}_i$.

If $\mathrm{dep}(\varphi) = 1$ then φ is generated by logical connectives from formulas of the form $K_i(\psi)$ where $\mathrm{dep}(\psi) = 0$, that is, ψ is generated by logical connectives from atomic formulas. Moreover, by Lemma 2 we can assume that ψ is non-trivial. It is easy to construct a model with a possible world at which $\neg K_i\psi \wedge \neg K_i\neg\psi$ holds true, that is, $\psi \in S_i$, and thus $\varphi \in \mathcal{F}_i^S$. Suppose we proved the claim for formulas of depth $\leq n$ and let $\mathrm{dep}(\varphi) = n+1$. Then φ is generated by logical connectives from formulas of the form $K_i(\psi)$ where $\mathrm{dep}(\psi) \leq n$. By Lemma 2 we can assume that each of the ψ's is non-trivial. Thus, it is enough to show that each non-trivial formula $K_i(\psi)$ with $\mathrm{dep}(\psi) \leq n$ is equivalent to a formula in $\mathcal{F}_i^S$. Now, if $\psi \in S_i$ then $K_i(\psi)$ is itself in $\mathcal{F}_i^S$ and we are done. Suppose then that $\psi \in R_i$. Then by Theorem 1, ψ is equivalent to a formula $\xi \in \mathcal{F}_i$ with $\mathrm{dep}(\xi) \leq \mathrm{dep}(\psi) \leq n$ and therefore $K_i\psi \equiv K_i\xi$. Since ψ is non-trivial, ξ is also non-trivial, and by the induction hypothesis, ξ is equivalent to a formula in $\mathcal{F}_i^S$. Since $\xi \in \mathcal{F}_i$, it follows by Theorem 1 that $\xi \in \mathrm{R_i}$ and thus by part (4) of proposition 1, $K_i\xi \equiv \xi$. We conclude that $K_i\psi$ is equivalent to ξ which is equivalent to a formula in $\mathcal{F}_i^S$. This implies that φ is equivalent to a formula in $\mathcal{F}_i^S$. ∎

Now, let $\Phi = \mathbb{K}_i^{\mathrm{S}} = \hat{\mathbb{K}}_i^{\mathrm{S}}$, and $\Psi = K_i\Phi \cup \neg K_i(S_i \setminus \Phi)$. As shown in part (b) of Proposition 3, Ψ is a consistent subset of $\mathbb{K}_i^{\mathrm{R}}$ and of $\hat{\mathbb{K}}_i^{\mathrm{R}}$. Moreover, Ψ is a maximal consistent subset of $\{K_i\varphi \mid \varphi \in S\}$. Since the formulas of $\mathcal{F}_i^S$ are generated from this set by logical connectives, it follows that there exists a unique maximal consistent subset of $\mathcal{F}_i^S$, Ψ', which contains Ψ. By Proposition 4, there exists a unique maximal consistent subset of the non-trivial formulas in $\mathrm{R_i}$ that contains Ψ'. By Theorem 1 any maximal consistent subset of $\mathrm{R_i}$ consists of a maximal consistent subset of the non-trivial formulas in $\mathrm{R_i}$ and all the theorems in $\mathrm{R_i}$. Thus, there exists a unique maximal consistent subset of $\mathrm{R_i}$ that contains Ψ. Since $\mathbb{K}_i^{\mathrm{R}}$ and $\hat{\mathbb{K}}_i^{\mathrm{R}}$ are maximal consistent subsets of $\mathrm{R_i}$ and each of them contains Ψ, it follows that $\mathbb{K}_i^{\mathrm{R}} = \hat{\mathbb{K}}_i^{\mathrm{R}}$. ∎

References

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

Hart, S. A. Hefetz and D. Samet(1996), "Knowing whether," "Knowing that," and The cardinality of state spaces, *Journal of Economic Theory*, 70, 249-256.

Hefetz, A. and D. Samet(1990), Hierarchies of knowledge: an unbounded stairway, *Mathematical Social Sciences*, 38, 157–170.

Fagin, R. and M. Y. Vardi,(1985), An internal semantics for modal logic, *Proceedings of the 17th ACM Symposium on Theory of Computing*, 305-315.

Fagin, R., J. Y. Halpern and M. Y.Vardi,(1991), A model-theoretic analysis of knowledge, *Journal of the ACM*, 38, 382-428.

Fagin, R., J. Y. Halpern and Y. Moses and M. Y.Vardi,(1995), *Reasoning About Knowledge*, The MIT Press, Cambridge, MA.

Hintikka, K. J. J.,(1962), *Knowledge and Belief: An Introduction to the Logic of the Two Notions*, Cornell University Press, Ithaca, NY.

Kripke, S. A.,(1980), *Naming and necessity*, Harvard University Press.

McGinn, C. (1975), A Priori and A Posteriori Knowledge, *Proceedings of the Aristotelian Society*, 76, 195-208.

Russell, B.(2020), A Priori Justification and Knowledge *The Stanford Encyclopedia of Philosophy*, Edward N. Z. (ed.), URL = https://plato.stanford.edu/archives/sum2020/entries/apriori.

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

Stalnaker, R.,(1996), "On Logics of Knowledge and Belief, *Philosophical Studies*, 128, 169-199.

Williamson, T. (2000), *The Philosophy of Philsophy*, Oxford University Press.