THE SURE-THING PRINCIPLE IN EPISTEMIC TERMS

DOV SAMET

ABSTRACT. Savage (1954) introduced the sure-thing principle in terms of the dependence of decisions on knowledge, but gave up on formalizing it in epistemic terms for lack of a formal definition of knowledge. Using a standard model of knowledge, the partition model, we examine the sure-thing principle, presenting two ways to capture it. One is in terms of knowledge operators, which we call the principle of *follow the knowledgeable*; the other is in terms of *kens*—bodies of agents' knowledge—which we call *independence of irrelevant knowledge*. We show that the two principles are equivalent. We present a stronger version of the independence of irrelevant knowledge and show that it is equivalent to the impossibility of agreeing to disagree on the decision made by agents, namely the impossibility of different decisions made by agents being common knowledge.

1. INTRODUCTION

1.1. **An example of the sure-thing principle.** The *sure-thing principle* (STP) was introduced by Savage (1954) using the following story.

> A businessman contemplates buying a certain piece of property. He considers the outcome of the next presidential election relevant. So, to clarify the matter to himself, he asks whether he would buy if he *knew* that the Democratic candidate were going to win, and decides that he would. Similarly, he considers whether he would buy if he *knew* that the Republican candidate were going to win, and again finds that he would. Seeing that he would buy in either event, he decides that he should buy, even though he does not *know* which event obtains, or will obtain, as we would ordinarily say. It is all too seldom that a decision can be arrived at on the basis of this principle, but except possibly for the assumption of simple ordering, I know of no other extralogical principle governing decisions that finds such ready acceptance. (emphasis added)

This story is told in terms of knowledge and decisions. However, when it came to describing the STP formally, Savage had to give up the epistemic nature of the story, namely the reference to agent's knowledge, for the following reason:

> The sure-thing principle cannot appropriately be accepted as a postulate in the sense that P1 is, because *it would introduce new undefined technical terms referring to knowledge and possibility that would render it mathematically useless without still more postulates governing these terms.* It will be preferable to regard the principle as a loose one that suggests certain formal postulates well articulated with P1. (emphasis added)

Thus, Savage did not consider his second postulate, P2, to be a formalization of the STP. Such formalization requires, as he says, a specification of the postulates governing the terms knowledge and possibility, that is, a formal model of knowledge which was not available at the time. Almost a decade after Savage's *The foundations of statistics*, Hintikka (1962) introduced formal modeling of knowledge, syntactic and semantic, in his *Knowledge and belief*, while a semantic multi-agent model of knowledge was introduced more than two decades later in Aumann (1976).

Here, we use a standard formal model of multi-agent knowledge for the formulation of the sure-thing principle that reflects its original intended meaning. We provide two formalizations of the STP and show their equivalence. We further show that the impossibility of agreeing to disagree on decisions is equivalent to a stronger version of the STP.

1.2. **A temporal model.** The simplest way to read Savage's story concerns an agent in two periods. In the second, the agent will be more knowledgeable, as he will know whether the Democrat candidate won the election ("D") or the Republican ("R"). As Savage puts it, "he considers the outcome of the next presidential election relevant" to the question of buying a certain property. Since the agent knows in the first period that being more knowledgeable in the second he will buy the property in either case, he should buy it in the first period at a time that he does not yet know who will win. The following is a rough draft of the STP that we call *follow the knowledgeable* (FTK):

> **FTK**: If an agent knows today her decision tomorrow, when she is more knowledgeable, then this should be her decision today.[1]

The following formulation of the STP emphasizes the dependence of the decision on the body of knowledge. If tomorrow the Democrat wins then our agent will know "D", and a fortiori "D or R". If tomorrow the Republican wins, the agent will know "R", and obviously "D or R". However, today, she knows only the intersection of these two bodies of knowledge, namely, "D or R". Now, the businessman in Savage's story "considers the outcome of the next presidential election relevant" to his decision. But to be sure, neither "D" nor "R" is relevant to his decision. Indeed, the very essence of the STP demonstrated in this story is that "R" and "D" prove to be *irrelevant* to his decision to buy the property. What remains relevant is just the knowledge common to both cases, namely "D or R", which is what the agent knows today.

In order to suggest a draft of this formulation of the STP we use the term *ken* to describe the body of knowledge of an agent. We further assume, along with Savage's assumption, that the decision of an agent depends on her knowledge, and more precisely on her ken. We call this version of the STP the *independence of irrelevant knowledge* (IIK).

> **IIK**: If the agent's decision is the same for all of her kens tomorrow, then what is relevant for her decision is the intersection of all these kens, namely her ken today, and therefore this should be her decision today.[2]

[1]One should not confuse this principle with the good advice "never put off until tomorrow what you can do today". The STP does not assume any gains or losses from the timing of the decision itself.

[2]IIK has features in common with various versions of the independence of irrelevant alternatives. It also resembles an ancient legal syllogism from the mishna called "binyan av" (prototype)

1.3. **A two-agent model.** In the temporal model the agent is split into two knowers, one for each period. Certain implicit assumptions are naturally made in this set up concerning the relation between the different knowers associated with the same agent, namely, that the agent is more knowledgeable in later periods, and, moreover, that the agent knows that.[3]

Instead of using a temporal model, we reformulate the FTK and IIK in a two-agent atemporal model. One can always think of the two agents as being the different knowers associated with a single agent at different times. The two-agent model forces us to spell out explicitly the relations implicitly assumed between the different knowers of the same agent. The assumption made before, concerning the agent in the temporal model, is translated in the two-agent model into the assumption that one agent is more knowledgeable than the other, and that the less knowledgeable agent knows it.

We can now state the two-agent versions of the STP where this assumption is made explicitly. One can think of Adam, in the following formulation, as being the agent today and Eve as the same agent tomorrow, or, alternatively, we can think of them as two different individuals who share the same interests.

The first formulation of the STP, in the two-agent set up is:

> **FTK** (two-agent version): If Adam knows that Eve is more knowledgeable, and he also knows Eve's decision, then this must be his decision too.

The second formulation is:

> **IIK** (two-agent version): If Adam's ken is the intersection of some of Eve's kens, and Eve's decision is the same for all these kens, then what is relevant for her decision is the intersection of all these kens, namely, Adam's ken, and therefore this should also be Adam's decision.

Our first result states:

> FTK *and the* IIK *are equivalent.*

1.4. **Strengthening the STP.** Adam and Eve play asymmetric roles in IIK. While there is a *single* ken for Adam, there is a *family* of Eve's kens. But the same reasoning can be applied to cases where each of them is endowed with a family of kens. We assume that if within a family of kens of an agent the decision is the same for all kens, then what is relevant for making this decision is the intersection of kens in the family. This implies the following principle.

> **Strong IIK**: If within a family of Adam's kens Adam's decision is the same for all kens, and within a family of Eve's kens Eve's decision is the same for all kens, and if moreover, the intersection of the kens in Adam's family of kens is the same as the intersection

---

or "hatsad hashaveh" (the common characteristic): If X and Y are two sets of circumstances, and in each the ruling is R, then the circumstances relevant to this ruling are those in $X \cap Y$, and the ruling R also applies to any set of circumstances that includes this intersection. (See an instance of this syllogism in the Babylonian Talmud, Tractate Sanhedrin 64b.) IIK is more restricted. It applies only to a ken which is the intersection of other kens, and not to kens that contain this ken.

[3]When we say that the agent is *more* knowledgeable in later periods, we are a little bit sloppy for the sake of brevity. We mean that in later periods he is *at least* as knowledgable.

of kens in Eve's family of kens, then Adam and Eve make the same decision.

Obviously, strong IIK implies IIK and bears a strong resemblance to it. In contrast, FTK does not lend itself to a straightforward strengthening. The "right" formulation, the one that is equivalent to strong IIK, requires the introduction of common knowledge. The condition in this case is well known from the agreement theorem and its generalizations to the non-probabilistic case (Aumann (1976), Samet (2007)): *the impossibility of agreeing to disagree* (IAD).

IAD: If Adam's and Eve's decisions are common knowledge between them, then these decisions are the same.

We show that

Strong IIK *is equivalent to* IAD.

We note that despite the similarity between strong IIK and IIK, there is a crucial difference between them. The two agents in IIK can be viewed as the same agent in different periods. The asymmetry between the two agents in the IIK is essential for this interpretation and reflects the asymmetry of time with respect to knowledge. No such interpretation of the two agent is possible for the strong IIK.

1.5. **The multi-agent case.** Strong IIK is formulated for two agents, but it can be formulated for a group of agents of any size. Suppose a family of kens is given for each agent in the group and within each such family the decision is the same for all the kens in the family. We can require in this case, in the spirit of IIK, that if the intersection of the kens in each family is the same for all agents, then the same decision is made in each family. However, if we require that strong IIK holds for each pair of agents in the group, then it must hold for the whole group, as any pair must make the same decision in their family of kens.

Similarly, IAD is formulated for a pair of agents, but can be formulated for a any group of agents by requiring that if their decisions are common knowledge between them, then the decisions must be the same. Again, if we require IAD for any pair in the group, then it holds for the whole group. Indeed, if the decisions of the agents are common knowledge between them, then the decision of any pair of agents is common knowledge between this pair of agents, and hence are the same. For these reasons, we formulate our results in a model two agents only.

2. Models of knowledge and decisions

Since the various definitions of the STP and their extensions involve only two agents we consider models of knowledge for two agents denoted 1 and 2. A *knowledge model* (Ω, π_1, π_2) consists of a set of *states* Ω, and for each agent $i \in \{1, 2\}$, a partition π_i of Ω. For each ω and i we denote by $\pi_i(\omega)$ the element of π_i that contains ω. Subsets of Ω are called *events*. We say that *i knows event E* at ω if $\pi_i(\omega) \subseteq E$. The event that *$i$ knows E*, denoted $K_i(E)$ is the set of all states in which i knows E. Thus, $K_i(E) = \{\omega \mid \ \pi_i(\omega) \subseteq E\}$. The function $K_i : 2^\Omega \to 2^\Omega$, thus defined is called i's *knowledge operator*.

Denote for each E, $K(E) = K_1(E) \cap K_2(E)$. The operator C defined by $C(E) = \cap_{m=1}^{\infty} K^m(E)$ is the *common knowledge* operator.

The set of all the events E that i knows at ω, that is, the set $\{E \mid \omega \in K_i(E)\}$, is called i's *ken* at ω and is denoted by $\mathrm{ken}_i(\omega)$.[4] Obviously, $\mathrm{ken}_i(\omega)$ consists of all the supersets of $\pi_i(\omega)$. We denote by Ken_i the family of all of i's kens, that is, $\mathrm{Ken}_i = \{\mathrm{ken}_i(\omega) \mid \omega \in \Omega\}$.

Let D be a nonempty set of *decisions*. A *decision function* $\mathbf{d}_i$ for agent i associates a decision with each of i's kens. That is, $\mathbf{d}_i$ is a function $\mathbf{d}_i \colon \mathrm{Ken}_i \to D$. A pair of decision functions for the two agents, $\mathbf{d} = (\mathbf{d}_1, \mathbf{d}_2)$ is called a *decision function profile*. With some abuse of notation we write $\mathbf{d}_i(\omega)$ for $\mathbf{d}_i(\mathrm{ken}_i(\omega))$. We denote by $[\mathbf{d}_i = d]$ the event that i's decision is d, namely $[\mathbf{d}_i = d] = \{\omega \mid \mathbf{d}_i(\omega) = d\}$.

3. The equivalence of FTK and IIK

The principle of follow the knowledgeable, which we now formally define, spells out in precise terms of models of knowledge and decision functions the verbal description of this principle as given in subsection 1.3. We note that $K_i\big(\neg K_i(E) \cup K_j(E)\big)$ is the event that i knows that if she knows the event E so does j. Thus, $\bigcap_{E \subseteq \Omega} K_i\big(\neg K_i(E) \cup K_j(E)\big)$ is the event that i knows that j is at least as knowledgeable. Our first formulation of the STP requires that if i knows that j is at least as knowledgable, and happens to know j's decision, than this should also be i's decision.

Follow the knowledgeable (FTK):

A decision function profile $\mathbf{d}$ *satisfies* FTK *if for agents* i, j, *and decision* d,

$$\bigcap_{E \subseteq \Omega} K_i\big(\neg K_i(E) \cup K_j(E)\big) \cap K_i([\mathbf{d}_j = d]) \subseteq [\mathbf{d}_i = d].$$

The independence of irrelevant knowledge below is a formal rendering of the description given to it in subsection 1.3. It says that if j's decision is d at each ken in a family of her kens $\mathcal{K}_j$, and if i's ken is the intersection of j's kens in $\mathcal{K}_j$, then i's decision is d. More formally,

Independence of irrelevant knowledge (IIK):

A decision function profile $\mathbf{d}$ *satisfies* IIK *when, for agents* i, j, *decisions* d_i, d_j $\mathbb{K}_i \in \mathrm{Ken}_i$, *and* $\mathcal{K}_j \subseteq \mathrm{Ken}_j$,

if

(1) $\mathbb{K}_i = \cap_{\mathbb{K}_j \in \mathcal{K}_j} \mathbb{K}_j$, *and*
(2) for each $\mathbb{K}_j \in \mathcal{K}_j$, $\mathbf{d}_j(\mathbb{K}_j) = d_j$, *and*
(3) $\mathbf{d}_i(\mathbb{K}_i) = d_i$

then $d_i = d_j$.

Theorem 1. *A pair of decision functions of the two agents satisfies the principle of follow the knowledgeable if and only if it satisfies independence of irrelevant knowledge.*

[4]Kens were introduced in Samet (1990) for more abstract models of knowledge. Using formal epistemic language, a ken of agent i is defined as a maximal set of sentences, Φ, such that the set of sentences $\{i \text{ knows } f | f \in \Phi\}$ is consistent.

4. Strengthening the STP

The following is an extension of IIK in which 1 and 2 play a symmetric role. Both are endowed with a family of kens: $\mathcal{K}_1 \subseteq \mathrm{Ken}_1$ and $\mathcal{K}_2 \subseteq \mathrm{Ken}_2$. If in each family the same decision is associated with each ken, then the relevant knowledge for making this decision is the intersection of the kens in each family. This intersection may fail to be a ken of an agent in the model. Yet, if the intersection of the kens in $\mathcal{K}_i$ coincides with the intersection of kens in $\mathcal{K}_j$ the same decision should be made by the two agents. The IIK is a special case of the SIIK, where the family of kens of one of the agents is a singleton.

Strong independence of irrelevant knowledge (SIIK):
A decision function profile $\mathbf{d}$ *satisfies* SIIK *when for agents* i, j, *decisions* d_i, d_j, *and non-empty sets of kens,* $\mathcal{K}_i \subseteq \mathrm{Ken}_i$, *and* $\mathcal{K}_i \subseteq \mathrm{Ken}_i$,

if

(1) $\cap_{\mathbb{K}_i \in \mathcal{K}_i} \mathbb{K}_i = \cap_{\mathbb{K}_j \in \mathcal{K}_j} \mathbb{K}_j$, *and*
(2) for each $\mathbb{K}_j \in \mathcal{K}_j$, $\mathbf{d}_j(\mathbb{K}_j) = d_j$, *and*
(3) for each $\mathbb{K}_i \in \mathcal{K}_i$, $\mathbf{d}_i(\mathbb{K}_i) = d_i$,

then, $d_i = d_j$.

We next formulate the condition of *impossibility of agreeing to disagree* (IAD), which is equivalent to SIIK, but is expressed in terms of knowledge operators rather in terms of kens. The condition requires that the agents cannot have common knowledge of their decisions when the decisions are not the same. It is analogous to a condition for models of knowledge and probabilistic beliefs, that agents cannot have common knowledge of their posterior probability of a given event when the two posterior probabilities differ. For the probabilistic case, Aumann (1976) showed that IAD holds when the posterior probabilities are derived from a prior which is common to both agents. Here we show that IAD is equivalent to SIIK which is a strong version of the STP.

Impossibility of agreeing to disagree (IAD):
A pair of decision functions $(\mathbf{d}_1, \mathbf{d}_2)$ *satisfies* IAD *if for* $d_1 \neq d_2$,

$$C\big([\mathbf{d}_1 = d_1] \cap [\mathbf{d}_2 = d_2]\big) = \emptyset.$$

Theorem 2. *A pair of decision functions of the two agents satisfies strong independence of irrelevant knowledge if and only if it satisfies the impossibility of agreeing to disagree.*

5. Discussion

The first attempt to use the sure-thing principle in an epistemic setup was made, independently, by Cave (1983) and Bacharach (1985), although it was the latter who used the term STP in this context. Both papers proposed a sufficient condition for IAD, generalizing the probabilistic agreement theorem of Aumann (1976) to the non-probabilistic cases, where at each state of the world a *decision* of each agent is specified, rather than a *posterior probability*.

Both papers use a partition model with a *virtual decision function* δ (a term suggested in Samet (2007)) from which individual decisions are derived. Such a function assigns a decision to *each* event. The interpretation is that the decision

$\delta(E)$ associated with an event E is the decision made when knowledge is given by E. This is very much in the spirit of Savage's example and the approach adopted here. The sure-thing principle in this setup says that for two disjoint events E and F for which $\delta(E) = \delta(F)$, it is the case that $\delta(E \cup F) = \delta(E)$.

Virtual decision functions are hard to interpret properly. Considering events which are not elements of the partition as describing knowledge is incongruent with the knowledge structure given by the partition. Moreover, by its very essence the STP cannot be applied to a single knower. The union $E \cup F$ purports to represent a body of knowledge—a ken—which is the intersection of the kens given by E and by F. But this idea is inconsistent with partition models: it is impossible for an agent in a partition model to have kens with an intersection that is also a ken of the *same agent*, except for the trivial case that the intersecting kens are identical. The only way to express the STP is through either the knowledge of an agent in *two* periods, or alternatively, as is the case here, the knowledge of *different* agents. Moses and Nachum (1990) were the first to study conceptual difficulties regarding virtual decision functions.

Samet (2007) found a sufficient condition for IAD in terms of the STP, avoiding the conceptual pitfalls in Cave (1983) and Bacharach (1985). For this purpose, Samet (2007) formulated STP as FTK under the name *interpersonal sure-thing principle* (ISTP). However, as is shown here, FTK, which is equivalent to IIK, is weaker than strong IIK, and hence weaker than IAD. In order to find a sufficient condition for IAD in terms of FTK, Samet (2007) had to fortify FTK. The condition that implies IAD is *ISTP-expandability* which requires that FTK also holds when another agent is added to the model. Here we show that IAD is not only *implied* by some version of the STP, but it *is* a strengthened version of STP, namely, strong IIK.

6. Proofs

A ken $\mathbb{K}_j$ of j is the set of all supersets of some element in π_j. Denote this element by $\pi_j(\mathbb{K}_j)$. The following is a straightforward set theoretic claim.

Claim 1. *For any j and $\mathcal{K}_j \subseteq \mathrm{Ken}_j$, $\cap_{\mathbb{K}_j \in \mathcal{K}_j} \mathbb{K}_j$ consists of all the supersets of $\cup_{\mathbb{K}_j \in \mathcal{K}_j} \pi_j(\mathbb{K}_j)$.*

Proposition 1. *For each ω, i and j, there exists a family of j's kens $\mathcal{K}_j \subseteq \mathrm{Ken}_j$ for which $\mathrm{ken}_i(\omega) = \cap_{\mathbb{K}_j \in \mathcal{K}_j} \mathbb{K}_j$, if and only if for each $\omega' \in \pi_i(\omega)$, $\pi_j(\omega') \subseteq \pi_i(\omega)$. Moreover, in this case $\mathcal{K}_j = \{\mathrm{ken}_j(\omega') \mid \omega' \in \pi_i(\omega)\}$.*

Proof: Suppose that for some $\mathcal{K}_j \subseteq \mathrm{Ken}_j$, $\mathrm{ken}_i(\omega) = \cap_{\mathbb{K}_j \in \mathcal{K}_j} \mathbb{K}_j$. By claim **??** the intersection on the right is the family of supersets of $\cup_{\mathbb{K}_j \in \mathcal{K}_j} \pi_j(\mathbb{K}_j)$. By the equality, this family coincides with the family of supersets of $\pi_i(\omega)$. Thus, $\pi_i(\omega) = \cup_{\mathbb{K}_j \in \mathcal{K}_j} \pi_j(\mathbb{K}_j)$. If $\omega' \in \pi_i(\omega)$, then for some $\mathbb{K}_j \in \mathcal{K}_j$, $\omega' \in \pi_j(\mathbb{K}_j)$. Therefore, $\pi_j(\omega') = \pi_j(\mathbb{K}_j)$ and hence $\pi_j(\omega') \subseteq \pi_i(\omega)$. We also conclude that $\mathbb{K}_j \in \mathcal{K}_j$ if and only if $\mathbb{K}_j = \mathrm{ken}_i(\omega')$ for some $\omega' \in \pi_i(\omega)$.

Conversely, suppose that for each $\omega' \in \pi_i(\omega)$, $\pi_j(\omega') \subseteq \pi_i(\omega)$. Let $\mathcal{K}_j = \{\mathrm{ken}_j(\omega') \mid \omega' \in \pi_i(\omega)\}$. By claim 1, $\cap_{\mathbb{K}_j \in \mathcal{K}_j} \mathbb{K}_j = \cup_{\omega' \in \pi_i(\omega)} \pi_j(\omega')$. Per our assumption, $\cup_{\omega' \in \pi_i(\omega)} \pi_j(\omega') \subseteq \pi_i(\omega)$. Since for each $\omega' \in \pi_i(\omega)$, $\omega' \in \pi_j(\omega')$, it follows that $\cup_{\omega' \in \pi_i(\omega)} \pi_j(\omega') = \pi_i(\omega)$. This means that $\mathrm{ken}_i(\omega) = \cap_{\mathbb{K}_j \in \mathcal{K}_j} \mathbb{K}_j$ ∎

Proposition 2. *For each ω, i and j, $\omega \in \cap_{E \subseteq \Omega} K_i\big(\neg K_i(E) \cup K_j(E)\big)$ if and only if for each $\omega' \in \pi_i(\omega)$, $\pi_j(\omega') \subseteq \pi_i(\omega)$.*

Proof: Suppose $\omega \in \cap_{E\subseteq\Omega} K_i\big(\neg K_i(E) \cup K_j(E)\big)$. Then for each E, $\pi_i(\omega) \subseteq \neg K_i(E) \cup K_j(E)$, and in particular $\pi_i(\omega) \subseteq \neg K_i(\pi_i(\omega)) \cup K_j(\pi_i(\omega))$. But, $\pi_i(\omega) = K_i(\pi_i(\omega))$, and therefore, $\pi_i(\omega) \subseteq K_j(\pi_i(\omega))$. Thus for each $\omega' \in \pi_i(\omega)$, $\omega' \in K_j(\pi_i(\omega))$, which means that $\pi_j(\omega') \subseteq \pi_i(\omega)$.

Conversely, suppose that for each $\omega' \in \pi_i(\omega)$, $\pi_j(\omega') \subseteq \pi_i(\omega)$. If $\omega \in K_i(E)$, then $\pi_i(\omega) \subseteq E$ and thus for each $\omega' \in \pi_i(\omega)$, $\omega' \in K_j(E)$. Hence, $\pi_i(\omega) \subseteq K_j(E)$. Thus, for each E, $\pi_i(\omega) \subseteq \neg K_i(E) \cup K_j(E)$. That is, $\omega \in K_i\big(\neg K_i(E) \cup K_j(E)\big)$. ∎

Proof of Theorem 1: Suppose that $(\mathbf{d}_1, \mathbf{d}_2)$ satisfies IIK. Let $\omega \in \cap_{E\subseteq\Omega} K_i\big(\neg K_i(E) \cup K_j(E)\big) \cap K_i([\mathbf{d}_j = d])$. By Propositions 1 and 2, $\mathrm{ken}_i(\omega) = \cap_{\omega'\in\pi_i(\omega)} \mathrm{ken}_j(\omega')$. Since $\omega \in K_i([\mathbf{d}_j = d])$, $\pi_i(\omega) \subseteq [\mathbf{d}_j = d]$. That is, for each $\omega' \in \pi_i(\omega)$, $\mathbf{d}_j(\mathrm{ken}_j(\omega')) = \mathbf{d}_j(\omega') = d$. Thus, by IIK, $\mathbf{d}_i(\mathrm{ken}_i(\omega)) = d$, i.e., $\omega \in [\mathbf{d}_i = d]$.

Conversely, suppose that $(\mathbf{d}_1, \mathbf{d}_2)$ satisfies FTK. Assume that for some $\mathrm{ken}_i(\omega) \in \mathrm{Ken}_i$, and $\mathcal{K}_j \subseteq \mathrm{Ken}_i$, $\mathrm{ken}_i(\omega) = \cap_{\mathbb{K}_j\in\mathcal{K}_j} \mathbb{K}_j$, and $\mathbf{d}_j(\mathbb{K}_j) = d$ for each $\mathbb{K}_j \in \mathcal{K}_j$. By Propositions 1 and 2, $\mathcal{K}_j = \{\mathrm{ken}_j(\omega') \mid \omega' \in \pi_i(\omega)\}$, and $\omega \in \cap_{E\subseteq\Omega} K_i\big(\neg K_i(E) \cup K_j(E)\big)$. As $\mathbf{d}_j(\omega') = \mathbf{d}_j(\mathrm{ken}_j(\omega')) = d$ for each $\omega' \in \pi_i(\omega)$, it follows that $\omega \in K_i([\mathbf{d}_j = d])$. By FTK, $\omega \in [\mathbf{d}_i = d]$, which means that $\mathbf{d}_i(\mathrm{ken}_i(\omega)) = d$. ∎

Proposition 3. *Let $\mathcal{K}_i \subseteq \mathrm{Ken}_i$, for $i = 1, 2$. Then, $\cap_{\mathbb{K}_1\in\mathcal{K}_1} \mathbb{K}_1 = \cap_{\mathbb{K}_2\in\mathcal{K}_2} \mathbb{K}_2$, if and only if there exist an event E such that $\mathcal{K}_i = \{\mathrm{ken}_i(\omega) \mid \omega \in E\}$, for $i = 1, 2$, and $C(E) = E$.*

Proof: We recall that the meet of the partitions π_1 and π_2 is the finest partition which is coarser than π_1 and π_1, denoted by $\pi_1 \wedge \pi_2$. An event E is the union of elements of π_1 and of elements of π_2, if and only if it is the union of elements of $\pi_1 \wedge \pi_2$. The knowledge operator defined by the partition $\pi_1 \wedge \pi_2$ is the common knowledge operator C. Hence, $C(E) = E$ if and only if E is the union of elements of $\pi_1 \wedge \pi_2$, or equivalently, if and only if it is the union of elements of π_i for $i = 1, 2$.

By Claim 1, the equality in the proposition holds if and only if $\cup_{\mathbb{K}_i\in\mathcal{K}_i} \pi_i(\mathbb{K}_i) = \cup_{\mathbb{K}_j\in\mathcal{K}_j} \pi_j(\mathbb{K}_j)$. Denote by E this event. Then, $\mathcal{K}_i = \{\mathrm{ken}_i(\omega) \mid \omega \in E\}$ for $i = 1, 2$. Moreover, for $i = 1, 2$, E is the union of elements of π_i, and hence it is a union of elements of the meet. Hence, $C(E) = E$. Conversely, if $C(E) = E$, and $\mathcal{K}_i = \{\mathrm{ken}_i(\omega) \mid \omega \in E\}$, for $i = 1, 2$, then, being a union of elements of π_i, $E = \cup_{\mathbb{K}_i\in\mathcal{K}_i} \pi_i(\mathbb{K}_i)$ for $i = 1, 2$. ∎

Proof of Theorem 2: Since C is a knowledge operator it has the following properties. It satisfies for each event X, $C^2(X) = C(X)$ (positive introspection), $C(X) \subseteq X$ (truth axiom), and for $X \subseteq Y$, $C(X) \subseteq C(Y)$ (monotonicity). We use these properties in the sequel.

Suppose that SKII holds, and assume that for $d_1 \neq d_2$, $C\big([\mathbf{d}_1 = d_1] \cap [\mathbf{d}_2 = d_2]\big) = E$ is a non-empty event contrary to IAD. Applying C to both sides of the equality, we conclude by positive introspection that $C(E) = E$. Since E is not empty, it follows that $\mathcal{K}_i = \{\mathrm{ken}_i(\omega) \mid \omega \in E\}$, for $i = 1, 2$, are not empty. Hence, by Proposition (3), condition 1 in the definition of SIIK holds. Moreover, by the truth axiom, $E \subseteq [\mathbf{d}_1 = d_1] \cap [\mathbf{d}_2 = d_2]$. Hence, condition 2 in the definition of SIIK holds, and thus, $d_1 = d_2$ contrary to our assumption.

Suppose, now that IAD holds and assume that d_i and $\mathcal{K}_i$, for $i = 1, 2$, satisfy the two conditions in the definition of SIIK. By the first condition, it follows from Proposition (3) that there exists a non-empty event E, such that $C(E) = E$, and such that $\mathcal{K}_i = \{\mathrm{ken}_i(\omega) \mid \omega \in E\}$ for $i = 1, 2$. From this, and condition 2 in the definition of SIIK, it follows that $E \subseteq [\mathbf{d}_1 = d_1] \cap [\mathbf{d}_2 = d_2]$. By the monotonicity of C, $E = C(E) \subseteq C([\mathbf{d}_1 = d_1] \cap [\mathbf{d}_2 = d_2])$. Since E is non-empty, it follows by IAD that $d_1 = d_2$. ∎

References

Aumann, R. J. (1976), Agreeing to disagree, *Ann. Statist.*, 4(6), 1236–1239.

Bacharach, M. (1985), Some extensions of a claim of Aumann in an axiomatic model of knowledge, *J. Econom. Theory*, 37(1), 167–190.

Cave, J. (1983), Learning to agree, *Economics Letters*, 12, 147–152.

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

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

Hintikka, J. (1962), *Knowledge and Belief*. Cornell University Press, Ithaca, NY.

Moses, Y. and G. Nachum (1990), Agreeing to disagree after all (extended abstract), In *Theoretical aspects of reasoning about knowledge (Pacific Grove, CA,)*, Morgan Kaufmann Ser. Represent. Reason., 151–168. Morgan Kaufmann, San Mateo, CA.

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

Samet, D. (2007), Agreeing to disagree: The non-probabilistic case, forthcoming *GEB*.

Savage, L. J. (1954), *The foundations of statistics.* John Wiley & Sons Inc., New York.