COMMON BELIEF OF RATIONALITY IN GAMES OF PERFECT INFORMATION

DOV SAMET

ABSTRACT. Aumann (1995) showed that for games with perfect information common knowledge of substantive rationality implies backward induction. Substantive rationality is defined in epistemic terms, that is in terms of knowledge. We show that when substantive rationality is defined in doxastic terms, that is, in terms of belief, then common *belief* of substantive rationality implies backward induction. Aumann (1998) showed that material rationality implies backward induction in the centipede game. This result does not hold when rationality is defined doxastically. However, if beliefs are interpersonally consistent then common belief of material rationality in the centipede game implies *common belief* of backward induction.

1. INTRODUCTION

Aumann (1995) proved that in perfect information games, common knowledge of substantive rationality implies the backward induction outcome. The language and model used in the paper are *epistemic* rather than *doxastic*, that is, they are formulated in terms of *knowledge* and not *belief*. The author clarified that the approach of the paper "does not work with probability 1 belief", and claimed that he could not fix problems with off-path behavior that are related to such belief. Aumann (1998) showed that in the centipede game, common knowledge of material rationality implies backward induction. Again, the author emphasized that the result was proved for knowledge and not for belief.

Here we reexamine these results in standard doxastic language and model. The model consists of a state space with probability functions associated with each state, describing the probabilistic beliefs of the players. The language has one doxastic operator for each player, called belief, interpreted as belief with probability 1. We describe the set theoretic structures required for representing such beliefs and characterize these structures axiomatically.

Our first observation is that the results of Aumann (1995) and (1998) cannot be stated with common belief instead of common knowledge. However, the notions of substantial and material rationality in these papers are epistemic ones, that is, they are formulated in terms of knowledge. Doxastic notions of rationality can be formulated in exactly the same way by changing knowledge to belief. Doing this we state and prove the doxastic analogue of Aumann (1995):

> Common belief of doxastic substantial rationality implies backward induction.

We are not as lucky with doxastic material rationality. As we show in Example 2 below, common belief of this rationality does not imply backward induction in the centipede game. A weaker version of this claim still holds when beliefs are interpersonally consistent. By this we mean that each player believes not only that her beliefs are correct (which follows from the axioms of belief) but also that all players' beliefs are correct. We state and prove:

> When beliefs are interpersonally consistent, common belief of doxastic material rationality implies common belief of backward induction.

The results here show that knowledge is not a necessary ingredient in the derivation of backward induction. The model of Samet (1996) shows that it is not a sufficient one: Common knowledge of rationality in this model does not imply backward induction.[1] Thus, the variance in conclusions reached in different models of perfect information games seems not to be related to the use of knowledge or belief. In accordance with previous findings, at least as game theoretic analysis is concerned, belief approximates knowledge and with appropriate care can substitute for it.[2]

2. Common knowledge of rationality

We use the same notations as Aumann (1995) and Aumann (1998). The set of player i's vertices is denoted by V_i, and the set of i's strategies is S_i. Knowledge is expressed in a standard partition model. The set of states is Ω. A *knowledge structure* is given by a set $(\Pi_i)_i$ of partitions of Ω . The knowledge operator K_i, associated with the partition Π_i, is defined by $\mathrm{K}_iE = \{\omega \mid \Pi_i(\omega) \subseteq E\}$, where $\Pi_i(\omega)$ is the element of Π_i that contains ω. The event that *all know* E is $\mathrm{K}E = \cap_i \mathrm{K}_i$. The event $\mathrm{CK}E$, that E is *common knowledge* is the event that all know E, all know that all know E and so on. Thus, $\mathrm{CK}E = \cap_{n \geq 1} \mathrm{K}^n E$. The n-tuple of the players' strategies at ω is $\boldsymbol{s}(\omega)$. We assume that each player knows her strategy. This means that $\boldsymbol{s}_i$ is measurable with respect to Π_i or alternatively, that

Knowing one's actual strategy. Each player knows she plays the strategy she actually plays. That is, ,

$$[\boldsymbol{s}_i = s_i] \subseteq \mathrm{K}_i[\boldsymbol{s}_i = s_i] \quad \text{for each } i \text{ and } s_i \in S_i. \tag{1}$$

Rationality of a player is defined in terms of her conditional payoff function at vertex v, $h_i^v(s)$. Since rationality is defined in Aumann (1995) and Aumann (1998) in terms of knowledge we label them epistemic.

Epistemic substantive rationality. Player i is *substantively rational epistemically* if she does not know of any strategy of hers that can increase her conditional payoff in any of her vertices. Thus the event that i is substantively rational epistemically is:

$$R_i^{\mathrm{es}} = \bigcap_{v \in V_i} \bigcap_{t_i \in S_i} \neg \mathrm{K}_i[h_i^v(\boldsymbol{s}; t_i) > h_i^v(\boldsymbol{s})]. \tag{2}$$

The event $R^{\mathrm{es}} = \cap_i R_i^{\mathrm{es}}$, is the event that all players are substantively rational epistemically.

---

[1] Ben-Porath (1997) characterized the strategy profiles that are consistent with common belief of rationality in perfect information games, and showed that it does *not* imply backward induction. It seems that Ben-Porath (1997) partially ascribes the different result of his model to the use of belief rather than knowledge. However, it is not difficult to modify Ben-Porath's model slightly in such a way that his characterization holds under common *knowledge* of rationality.

[2] See, e.g., Monderer and Samet (1989).

Epistemic material rationality. Player i is *materially rational epistemically* if for each of i's vertices v and strategy t_i, either i knows that v is not reached, or else, i does not know that t guarantees her a higher conditional payoff at v *when* v *is reached.* We denote by Ω^v the event that vertex v is reached. The event R_i^{em} that player i is materially rational epistemically is an exact rendering of this sentence to the language of the model.

$$R_i^{\mathrm{em}} = \bigcap_{v \in V_i} \bigcap_{t_i \in S_i} (\mathrm{K}_i \neg \Omega^v) \;\cup\; \neg \mathrm{K}_i(\neg \Omega^v \cup [h_i^v(\boldsymbol{s}; t_i) > h_i^v(\boldsymbol{s})]). \tag{3}$$

The event $R^{\mathrm{em}} = \cap_i R_i^{\mathrm{em}}$, is the event that all players are materially rational epistemically.[3]

We denote by I the event that the backward induction path is realized. Common knowledge of rationality has the following implications.

Theorem 1. [Aumann (1995)] For generic games, $\mathrm{CK}R^{\mathrm{es}} \subseteq I$.[4]

Theorem 2. [Aumann (1998)] For the centipede game, $\mathrm{CK}R^{\mathrm{em}} \subseteq I$.

3. BELIEFS

In order to examine the implications of common belief of rationality, we first present standard models of belief.[5]

3.1. Belief structures.

Probabilistic belief structures. A *probabilistic belief structure* on Ω is a set of *type functions* $(t_i)_i$ on Ω. For each i and ω, $t_i(\omega)$, the *type* of i at ω, is a probability function on Ω, representing i's beliefs at ω. Let Π_i be the partition of Ω into sets of states with the same values of t_i, that is, $\Pi_i(\omega) = \{\omega' \mid t_i(\omega') = t_i(\omega)\})$. We require that for each state ω, $t_i(\omega)(\Pi_i(\omega)) = 1$, which means that i is always certain of her type. For each i, $\mathrm{B}_i^1 E$ is the event that i is certain of E. That is, $\mathrm{B}_i^1 E = \{\omega \mid t_i(\omega)(E) = 1\}$.[6]

Since we are interested only in the belief operators B_i^1, which are defined in terms of events of probability 1, we can use a simpler structure which is induced by the probabilistic belief structure.

Belief structures. A *belief structure* on Ω is a set of pairs $((\Pi_i, b_i))_i$, where Π_i is a partition of Ω, and b_i is a function, $b_i \colon \Omega \to 2^\Omega \setminus \{\emptyset\}$, which is measurable with respect to Π_i and for each ω, $b_i(\omega) \subseteq \Pi_i(\omega)$. We associate with the belief structure belief operators B_i defined by $\mathrm{B}_i E = \{\omega \mid b_i(\omega) \subseteq E\}$.

---

[3] Aumann (1998) defined ex-post material rationality in terms of ex-post knowledge operators at vertex v, and proved Theorem 2 for common knowledge of this type of rationality. However, Samet (2011) showed that ex-post knowledge is not required for Aumman's definition. Moreover, the event that a player is materially rational, as defined here, is the event that the player knows that she is ex-post materially rational. In particular, common knowledge of material rationality and common knowledge of ex-post material rationality are one and the same event.

[4] It is not clear why Aumann (1995) does not state the stronger theorem, that common knowledge of substantive rationality implies the backward-induction *strategies*.

[5] For a comprehensive survey of models of belief and knowledge and their applications to game theory, see Battigalli and Bonanno (1999). Our presentation avoids the modal logic apparatus and uses set theoretic terminology instead.

[6] The operators B_i^1 is the 1-belief operator in the family of p-belief operators, B_i^p, studied in Monderer and Samet (1989).

Obviously, a probabilistic belief structure on Ω *induces* a belief structure on Ω, where Π_i is the partition of Ω into i's types, and $b_i(\omega)$ is the set of states in $\Pi_i(\omega)$ of positive $t_i(\omega)$ probability. Conversely, each belief structure on Ω is induced by some probabilistic belief structure on Ω.

Claim 1. *Let* $(\mathrm{B}_i^1)_i$ *be the operators associated with a probabilistic belief structure on* Ω*. The operators* $(\mathrm{B}_i)_i$ *that are associated with the induced belief structure, satisfy* $\mathrm{B}_i = \mathrm{B}_i^1$*.*

It is easy to see that each of the belief operators B_i associated with a belief structure satisfies the following four axioms.[7]

(K) $\mathrm{B}(\neg E \cup F) \cap \mathrm{B}E \subseteq \mathrm{B}F$
(D) $\mathrm{B}E \subseteq \neg\mathrm{B}\neg E$
(4) $\mathrm{B}E \subseteq \mathrm{B}\mathrm{B}E$
(5) $\neg\mathrm{B}E \subseteq \mathrm{B}\neg\mathrm{B}E$

Moreover, these axioms characterize belief structures as we state next.

Proposition 1. *A set of operators* $\mathrm{B}_i\colon 2^\Omega \to 2^\Omega$ *satisfy the axioms* KD45 *if and only if there exists a belief structure on* Ω *such that the belief operators associated with it are the operators* B_i*.*

The formulation of axiom K is standard in modal logic. However it is equivalent to the distributivity of the operator B over intersection, as we state next.

Proposition 2. *Axiom* K *holds if and only the following axiom of distributivity holds.*

(Dist) $\mathrm{B}(E \cap F) = \mathrm{B}E \cap \mathrm{B}F$

By plugging $E \subseteq F$ into axiom Dist, we deduce the monotonicity of B:

Claim 2. *Axiom Dist implies that for* $E \subseteq F$, $\mathrm{B}E \subseteq F$*.*

3.2. **Belief and knowledge.** Belief is one axiom short of knowledge. An operator K on Ω is a knowledge operator derived from a partition of Ω if and only if it satisfies the four axioms of belief and the *truth axiom*:[8]

(T) $\mathrm{K}E \subseteq E$

This axiom states that knowledge is correct. That is, if E is known it must be true. It can can also be written equivalently as $\neg\mathrm{K}E \cup E = \Omega$. For a belief operator B, $\neg\mathrm{B}E \cup E$ does not always have to hold (i.e., it does not have to be true in all states). However, it is always *believed* to be correct.

Claim 3. *For each* E*,* $\mathrm{B}(\neg\mathrm{B}E \cup E) = \Omega$*.*

Indeed, by the monotonicity of B, $\mathrm{B}(E) \subseteq \mathrm{B}(\neg\mathrm{B}(E) \cup E)$. By (5) and monotonicity $\neg\mathrm{B}(E) \subseteq \mathrm{B}(\neg\mathrm{B}(E)) \subseteq \mathrm{B}(\neg\mathrm{B}(E) \cup E)$. Thus, $\Omega \subseteq \mathrm{B}(\neg\mathrm{B}(E) \cup E)$.

---

[7] The names of these axioms are standard in modal logic. Hintikka (1962) suggested this logic as the logic of belief. As we see here, this is also the logic of certainty, or belief in probability 1.

[8] By adding the truth axiom we can omit axioms D and 4 which are derived from K, 5, and T.

The partitions Π_i of a belief structure define knowledge operators K_i.[9] These operators satisfy for each E,

(KB1) $\quad \mathrm{K}E \subseteq \mathrm{B}E$

(KB2) $\quad \mathrm{B}E \subseteq \mathrm{KB}E$

That is, knowledge implies belief and belief implies knowledge of the belief.[10]

The common belief operator is defined similarly to the common knowledge operator. We denote by $\mathrm{B}E$ the event that all know E, that is, $\mathrm{B}E = \cap_i \mathrm{B}_i E$, and by B^n, B to the power of n. The common belief operator is defined by $\mathrm{CB}E = \cap_{n\geq 1} \mathrm{B}^n E$. Since $\mathrm{K}_i E \subseteq \mathrm{B}_i E$, and as K_i and B_i are monotonic operators it follows that $\mathrm{CK}E \subseteq \mathrm{CB}E$. Thus, common belief is a weaker condition than common knowledge.

4. COMMON BELIEF OF RATIONALITY

Theorems 1 and 2 cannot be strengthened by replacing common knowledge of rationality with the larger event of common belief of rationality, as is shown in Example 1 below. In order to formulate doxastic versions of these theorems, we assume that our language has at its disposal only statements about belief and not about knowledge.

4.1. Common belief of substantive rationality.

We replacethe operator K_i with B_i in (1) and (2). Thus, the assumption of knowing one's strategy, in (1), becomes,

Believing one's actual strategy. Each player believes she plays the strategy she actually plays. That is, for each i,

$$[\boldsymbol{s}_i = s_i] \subseteq \mathrm{B}_i[\boldsymbol{s}_i = s_i] \quad \text{for each } i \text{ and } s_i \in S_i. \tag{4}$$

Doxastic substantive rationality is defined similarly to (2).

Doxastic substantive rationality: The event that i is *substantively rational doxastically* is:

$$R_i^{\mathrm{ds}} = \bigcap_{v\in V_i} \bigcap_{t_i \in S_i} \neg \mathrm{B}_i[h_i^v(\boldsymbol{s}; t_i) > h_i^v(\boldsymbol{s})], \tag{5}$$

As before, doxastic substantive rationality holds at $R^{\mathrm{ds}} = \cap_i R_i^{\mathrm{ds}}$.

By axiom KB1, doxastic substantive rationality is stronger than the epistemic version. That is, $R^{\mathrm{ds}} \subseteq R^{\mathrm{es}}$. It turns out that this strengthening compensates for the weakening of common knowledge, and the result is that Theorem 1 can be stated *mutatis mutandis* in its doxastic version.

Theorem 3. *For generic games,* $\mathrm{CB}R^{\mathrm{ds}} \subseteq I$.

---

[9] Each belief structure defines a unique knowledge operator that satisfies axiom KB1 and KB2 below. We refer to this as the *explicit definability* of knowledge in terms of belief. However, it is impossible to define the knowledge operator *explicitly* in terms of the belief operator. For an explanation of the difference between the two types of definability, see Halpern *et al.* (2009).

[10] If operators K_i satisfy these axioms then they are necessarily defined by the partitions Π_i. The first axiom guarantees that each partition Π_i is at least as fine as the partition associated with K_i and the second that it is at least as coarse.

4.2. **Common belief of material rationality.** The doxastic analogous of (3) is:

Doxastic material rationality: The event that i is *materially rational doxastically* is:

$$R_i^{\mathrm{dm}} = \bigcap_{v \in V_i} \bigcap_{t_i \in S_i} (\mathrm{B}_i \neg \Omega^v) \ \cup \ \neg \mathrm{B}_i(\neg \Omega^v \cup [h_i^v(\boldsymbol{s}; t_i) > h_i^v(\boldsymbol{s})]). \tag{6}$$

Doxastic material rationality is the event $R^{\mathrm{dm}} = \cap_i R_i^{\mathrm{dm}}$.

In contrast to the case of substantive rationality, the doxastic version of material rationality is not stronger than the epistemic one. It is possible that doxastic material rationality holds true by virtue of $\mathrm{B}_i \neg \Omega^v$ being true for some vertex v while $\neg \mathrm{B}_i(\neg \Omega^v \cup [h_i^v(\boldsymbol{s}; t_i) > h_i^v(\boldsymbol{s})])$, and a fortiori $\neg \mathrm{K}_i(\neg \Omega^v \cup [h_i^v(\boldsymbol{s}; t_i) > h_i^v(\boldsymbol{s})])$, are false. It is possible in this case, that $\mathrm{K}_i \neg \Omega^v$ is false, and as a result epistemic material rationality does not hold. Example 2 below demonstrates that a doxastic version of Theorem 2 is false. However, under the condition that beliefs are interpersonally consistent, which is defined below, we can state a weaker version of this theorem.

Theorem 4. *Assume that beliefs are interpersonally consistent. Then, for the centipede game,* $\mathrm{CB}R^{\mathrm{dm}} \subseteq \mathrm{CB}I$.

4.3. **Interpersonal consistency.** Claim 3 states that each player always believes that her own beliefs are correct. We say that beliefs are interpersonally consistent when each player believes that not only she, but all other players have correct beliefs. Formally, beliefs in a belief structure are *interpersonally consistent* if for each i, j and E:[11]

$$\mathrm{B}_i(\neg \mathrm{B}_j E \cup E) = \Omega \tag{Con}$$

Interpersonal consistency can be expressed in terms of perceived worlds. The *world perceived by* i is the minimal event F that satisfies $\mathrm{B}_i(F) = \Omega$. To justify this definition, consider the family $\mathcal{F}$ of all events F that satisfy the equality. By axioms D and Dist, $\mathcal{F} \neq \emptyset$ as $\neg \mathrm{B}_i \Omega \subseteq \mathrm{B}_i \neg \Omega = \emptyset$. By axiom Dist, $\cap_{F \in \mathcal{F}} F \in \mathcal{F}$ and by D it is not empty. This intersection is the world perceived by i.

Proposition 3. *Beliefs are interpersonally consistent if and only if all players perceive the same world.*

Interpersonal consistency of beliefs is related to interpersonal consistency of probabilistic beliefs. A probability $p \in \Delta(\Omega)$ is a *common prior* for a probabilistic belief structure $(t_i)_i$ on Ω with type partitions Π_i, if for each i and ω, $p(\Pi_i(\omega)) > 0$, and $t_i(\omega) = p(\cdot \mid \Pi_i(\omega))$. The types in a probabilistic belief structure are *equivalent* when the types of the players in each state are equivalent probability functions. That is, for each ω, i j, and E, if $t_i(\omega)(E) = 0$ then $t_j(\omega)(E) = 0$.[12]

Proposition 4. *The following three conditions are equivalent for the beliefs in a given belief structure.*

(1) *The beliefs in a belief structure are interpersonally consistent.*

---

[11] Bonanno and Nehring (1998) introduced this property under the more descriptive term "belief of no error". We adopted the term consistency as a hint to the relation with consistency of probabilistic beliefs (see Proposition 4).

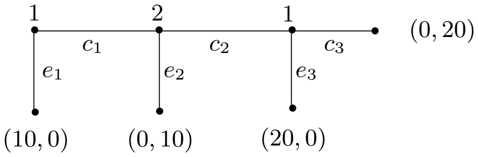
[12] See Bonanno and Nehring (1998) for an extensive discussion of the relation between Aumann's (1976) agreement theorem and interpersonal consistency of beliefs.

(2) *The belief structure is induced by a probabilistic belief structure that has a common prior.*
(3) *The belief structure is induced by a probabilistic belief structure in which the types are equivalent.*

5. EXAMPLES

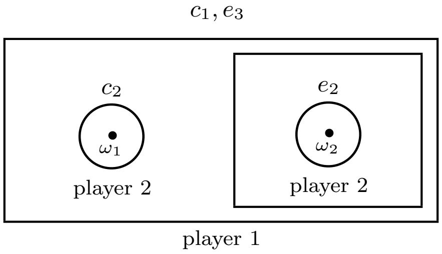
Example 1. The following simple example demonstrates that Theorems 1 and 2 cannot be strengthened by changing the common knowledge events $\mathrm{CK}R^{\mathrm{em}}$ and $\mathrm{CK}R^{\mathrm{es}}$ to the possibly larger common belief events $\mathrm{CB}R^{\mathrm{em}}$ and $\mathrm{CB}R^{\mathrm{es}}$.

Consider the three-legged centipede game in Figure 1 and its model in Figure 2. In the model, player 2's partition is given by the circles, and player 1's partition by the box. The inner box is what player 1 believes to be the state in both states, that is, $\mathrm{B}_1\{\omega_2\} = \Omega$. The players' strategies are written above each element of the partitions.



A three-legged centipede game

FIGURE 1



Theorems 1 and 2 fail for common belief

FIGURE 2

- Player 2 is substantively and materially rational epistemically at ω_2, as e_2 is a best response to (c_1, e_3).
- Player 1 does not know of any strategy that yields a higher payoff at the root (which is of course reached at both states) because (c_1, e_3) is a best response to c_2 which is played at ω_1. Her second vertex is not reached but even there, her strategy yields the best possible conditional payoff. Thus player 1 is substantively and materially rational epistemically in both states.

- Obviously, player 2 is neither substantively nor materially rational epistemically at ω_1.
- Thus, $R^{\text{em}} = R^{\text{es}} = \{\omega_2\}$.
- Since $\text{B}_1\{\omega_2\} \cap \text{B}_2\{\omega_2\} = \{\omega_2\}$ it follows that $\text{CB}\{\omega_2\} = \{\omega_2\}$ and therefore $\text{CB}R^{\text{em}} = \text{CB}R^{\text{es}} = \{\omega_2\}$.

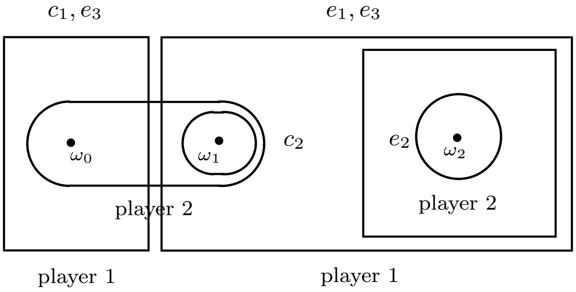
However at ω_2 the game is continued by the first player, and is terminated by the second, which is not the backward induction outcome.

Note that player 1 is substantively and materially rational epistemically, but not doxastically. At ω_2, e_1 yields player 1 a higher payoff than (c_1, e_3). Since player 1 believes $\{\omega_2\}$, she believes that the strategy c_1 will yield a higher payoff when her second vertex is reached.

Example 2. Figure 3 depicts a model of the game in Figure 1. The boxes are elements of the partition of player 1, the round figures—of player 2. The inner box and oval are the events believed by player 1 and 2, respectively. Obviously, players 1 and 2 are substantively and materially rational at ω_2, both epistemically and doxastically.

Player 2 is *not* substantively rational doxsastically at $\{\omega_0, \omega_1\}$. Indeed, she believes that player 1's strategy is (e_1, e_3), and thus her strategy is dominated by e_2 at her only vertex. Therefore, $\{\omega_2\}$ is the event of common belief of doxastic substantive rationality, and as implied by Theorem 3, the backward induction outcome holds in this state.

In contrast, player 2 *is* materially rational doxastically at $\{\omega_0, \omega_1\}$, since her vertex is not reached at ω_1, and therefore she believes that her vertex is not reached. Thus, common belief of doxastic material rationality holds everywhere. However, the backward induction outcome does not hold at ω_0. Moreover, there is not even common belief that this outcome holds.



The doxastic version of Theorem 2 fails.

FIGURE 3

6. PROOFS

Proof of Proposition 2. Suppose that for each E and F, $\text{B}(\neg E \cup F) \cap \text{B}E \subseteq \text{B}F$. For $E = F$ we have $\text{B}\Omega \cap \text{B}E \subseteq \text{B}E$. Thus for each E, $\text{B}E \subseteq \text{B}\Omega$. For $E \subseteq F$ we have $\text{B}\Omega \cap \text{B}E \subseteq \text{B}F$, thus, $\text{B}E \subseteq \text{B}F$, that is, B is monotonic. By monotonicity, for all events X and Y, $\text{B}(X \cap Y) \subseteq \text{B}X \cap \text{B}Y$. For the converse inclusion, we use monotonicity again to conclude that $\mathrm{B}X \cap \mathrm{B}Y \subseteq \mathrm{B}(\neg Y \cup (X \cap Y)) \cap \mathrm{B}Y \subseteq \mathrm{B}(X \cup Y)$. Suppose now that $\mathrm{B}(X \cap Y) = \mathrm{B}X \cap \mathrm{B}Y$. Substituting $X \subseteq Y$ we conclude that B is monotonic. Using the distributivity of B and its monotonicity, $\mathrm{B}(\neg E \cup F) \cap \mathrm{B}E = \mathrm{B}(F \cap E) \subseteq \mathrm{B}F$. $\square$

Proof of Proposition 1. It is enough to prove this proposition for a single belief operator which we denote by B. It is easy to check that a belief operator in a belief structure satisfies the axioms. We show the converse. Suppose that B satisfies the four axioms. Let $\beta(\omega) = \{E \mid \omega \in \mathrm{B}E\}$ be the set of the events believed at ω. Let Π be the partition of Ω into subsets of states with the same beliefs. That is, for each ω and ω', $\omega' \in \Pi(\omega)$ when $\beta(\omega) = \beta(\omega')$. Define for each ω, $c(\omega) = \cap_{E \in \beta(\omega)} E$. Obviously, if $\omega' \in \Pi(\omega)$ then $c(\omega') = c(\omega)$. We need to show that $c(\omega) \neq \emptyset$. By K,

$$\mathrm{B}\, c(\omega) = \cap_{E \in \beta(\omega)} \mathrm{B}E. \tag{7}$$

Since ω is in the right hand side, $\mathrm{B}\, c(\omega) \neq \emptyset$. But this implies that $c(\omega) \neq \emptyset$, since by K and D, $\mathrm{B}\emptyset = \mathrm{B}(E \cap \neg E) = \mathrm{B}E \cap \mathrm{B}\neg E = \emptyset$. Thus (Π, c) is a belief structure. To complete the proof, we show that B is the operator associated with (Π, c).

We first show that,

$$\Pi(\omega) = \cap_{E \in \beta(\omega)} \mathrm{B}E. \tag{8}$$

Obviously, for each $\mathrm{B}E$ such that $\omega \in \mathrm{B}E$ it also holds that $\Pi(\omega) \subseteq \mathrm{B}E$, which shows that $\Pi(\omega) \subseteq \cap_{E \in \beta(\omega)} \mathrm{B}E$. Conversely, let $\omega' \in \cap_{E \in \beta(\omega)} \mathrm{B}E$, then $\beta(\omega) \subseteq \beta(\omega')$. To show that there is equality, suppose to the contrary that for some E, $\omega' \in \mathrm{B}E$ but $\omega \notin \mathrm{B}E$. Then $\omega \in \neg\mathrm{B}E$, and by axiom 5, $\omega \in \mathrm{B}\neg\mathrm{B}E$. Thus, as $\neg\mathrm{B}E \in \beta(\omega)$ it follows that $\omega' \in \mathrm{B}\neg\mathrm{B}E$. On the other hand by axiom 4, $\omega' \in \mathrm{B}\mathrm{B}E$. This contradicts D. The equality $\beta(\omega) = \beta(\omega')$ implies that $\omega' \in \Pi(\omega)$, which completes the proof of (8).

We need to show that for each ω and E, $\omega \in \mathrm{B}E$ if and only if $c(\omega) \subseteq E$. Suppose $c(\omega) \subseteq E$. By monotonicity $\mathrm{B}\, c(\omega) \subseteq \mathrm{B}E$. By (7) and (8), $\Pi(\omega) \subseteq \mathrm{B}E$ and thus, $\omega \in \mathrm{B}(E)$. The converse implication follows immediately from the definition of c. $\square$

Lemma 1. *Suppose that $F \subseteq \mathrm{B}_i F$, then for each $\omega \in F$, $b_i(\omega) \subseteq \Pi_i(\omega) \cap F$. The set $((\hat{\Pi}_i, \hat{b}_i))_i$, where $\hat{\Pi}_i(\omega) = \Pi_i(\omega) \cap F$ and $\hat{b}_i$ is the restriction of b_i to F is a belief structure on $\hat{\Omega} = F$. The knowledge operators $\hat{\mathrm{K}}_i$ and belief operators $\hat{\mathrm{B}}_i$ associated with this belief structure satisfy for each event E in Ω:*

$$\hat{\mathrm{K}}_i(E \cap F) = \mathrm{K}_i(\neg F \cup E) \cap F, \tag{9}$$

$$\hat{\mathrm{B}}_i(E \cap F) = (\mathrm{B}_i E) \cap F, \tag{10}$$

$$\neg\hat{\mathrm{B}}_i(E \cap F) = \neg(\mathrm{B}_i E) \cap F, \tag{11}$$

where the complement on the left hand side of the last equality is with respect to F.

Proof. Since $F \subseteq \mathrm{B}_i F$, for each $\omega \in F$, $b_i(\omega) \subseteq F$. Thus, $b_i(\omega) \subseteq \Pi_i(\omega) \cap F = \hat{\Pi}_i(\omega)$. Hence, $((\hat{\Pi}_i, \hat{b}_i))_i$ is a belief structure on $\hat{\Omega}$.

Now, $\hat{\mathrm{K}}_i(E \cap F) = \{\omega \in F \mid \hat{\Pi}_i(\omega) \subseteq E \cap F\} = \{\omega \in F \mid \Pi_i(\omega) \cap F \subseteq E \cap F\} = \{\omega \mid \Pi_i(\omega) \subseteq \neg F \cup E\} \cap F = \mathrm{K}_i(\neg F \cup E) \cap F$, and $\hat{\mathrm{B}}_i(E \cap F) = \{\omega \in F \mid \hat{b}_i(\omega) \subseteq E \cap F\} = \{\omega \in F \mid b_i(\omega) \subseteq E \cap F\} = \{\omega \in F \mid b_i(\omega) \subseteq E\} = \{\omega \mid b_i(\omega) \subseteq E\} \cap F = (\mathrm{B}_i E) \cap F$. Thus, $F \setminus \hat{\mathrm{B}}_i(E \cap F) = F \cap \neg((\mathrm{B}_i E) \cap F) = \neg(\mathrm{B}_i E) \cap F$. $\square$

Lemma 2. *The assumption of* believing *one's actual strategy is the same as the assumption of* knowing *one's actual strategy.*

Proof. Condition (1) follows from (4) by axiom KB1. For the converse, observe that for $s_i \neq s_i'$, $[\boldsymbol{s}_i = s_i] \cap [\boldsymbol{s}_i = s_i'] = \emptyset$. By axioms K and D, $\mathrm{B}_i[\boldsymbol{s}_i = s_i] \cap \mathrm{B}_i[\boldsymbol{s}_i = s_i'] = \emptyset$. But $\cup_{s_i \in S_i}[\boldsymbol{s}_i = s_i] = \Omega$. Thus (4) implies that for each s_i, $[\boldsymbol{s}_i = s_i] = \mathrm{B}_i[\boldsymbol{s}_i = s_i]$. By axiom KB2, $\mathrm{B}_i[\boldsymbol{s}_i = s_i] \subseteq \mathrm{K}_i\mathrm{B}_i[\boldsymbol{s}_i = s_i]$, which by the previous equality is $\mathrm{K}_i[\boldsymbol{s}_i = s_i]$. ☐

Proof of Theorem 3. Let $F = \mathrm{CB}R_i^{\mathrm{ds}}$. By Proposition 3 in Monderer and Samet (1989),

$$F \subseteq \mathrm{B}_iF \cap \mathrm{B}_iR_i^{\mathrm{ds}}. \tag{12}$$

Thus, Lemma 1 applies to F. Denote by $\hat{\boldsymbol{s}}$ the restriction of $\boldsymbol{s}$ to $\hat{\Omega}$.

We show that assuming (4), (1) and (2) hold for the knowledge operators $\hat{\mathrm{K}}_i$.

By (4), Lemma 2, the monotonicity of K_i, and (9): $[\hat{\boldsymbol{s}}_i = s_i] = [\boldsymbol{s}_i = s_i] \cap F \subseteq \mathrm{K}_i[\boldsymbol{s}_i = s_i] \cap F \subseteq \mathrm{K}_i([\boldsymbol{s}_i = s_i] \cup \neg F) \cap F = \hat{\mathrm{K}}_i([\boldsymbol{s}_i = s_i] \cap F) = \hat{\mathrm{K}}_i([\hat{\boldsymbol{s}}_i = s_i])$. This shows that (1) holds.

By (11) and axiom KB1,

$$\begin{aligned} \neg\mathrm{B}_i[h_i^v(\boldsymbol{s};t_i) > h_i^v(\boldsymbol{s})] \cap F &= \neg\hat{\mathrm{B}}_i([h_i^v(\boldsymbol{s};t_i) > h_i^v(\boldsymbol{s})] \cap F) \\ &= \neg\hat{\mathrm{B}}_i[h_i^v(\hat{\boldsymbol{s}};t_i) > h_i^v(\hat{\boldsymbol{s}})] \\ &\subseteq \neg\hat{\mathrm{K}}_i[h_i^v(\hat{\boldsymbol{s}};t_i) > h_i^v(\hat{\boldsymbol{s}})] \end{aligned} \tag{13}$$

Thus, for each i,

$$R_i^{\mathrm{ds}} \cap F \subseteq \hat{R}_i^{\mathrm{es}} \tag{14}$$

where the last event with the hat is i's rationality as defined in (2) for $\hat{\mathrm{K}}_i$. Since K_i satisfies axioms K, 5, and T,

$$\hat{R}_i^{\mathrm{es}} = \hat{\mathrm{K}}_i\hat{R}_i^{\mathrm{es}} \tag{15}$$

Also for every pair of knowledge and belief operators K and B in a belief structure,

$$\mathrm{BK}E \subseteq \mathrm{K}E, \tag{16}$$

because by axioms 5 and KB1, $\mathrm{BK}E \cap \neg\mathrm{K}E = \mathrm{BK}E \cap \mathrm{K}\neg\mathrm{K}E \subseteq \mathrm{BK}E \cap \mathrm{B}\neg\mathrm{K}E$, which by axioms Dist and D is empty.

Now, by (12), (10), the monotonicity of B_i, and (14), $F = \mathrm{B}_i(R^{\mathrm{ds}}) \cap F = \hat{\mathrm{B}}_i(R^{\mathrm{ds}} \cap F) \subseteq \hat{\mathrm{B}}_i(R_i^{\mathrm{ds}} \cap F) \subseteq \hat{\mathrm{B}}_i\hat{R}_i^{\mathrm{es}}$. By (15) and (16), $\hat{\mathrm{B}}_i\hat{R}_i^{\mathrm{es}} = \hat{\mathrm{B}}_i\hat{\mathrm{K}}_i\hat{R}_i^{\mathrm{es}} \subseteq \hat{\mathrm{K}}_i\hat{R}_i^{\mathrm{es}} = \hat{R}_i^{\mathrm{es}}$. Thus, for each i, $F = \hat{R}_i^{\mathrm{es}}$, hence $F = \hat{R}^{\mathrm{es}}$ and therefore $F = \widehat{\mathrm{CK}}\hat{R}^{\mathrm{es}}$. By Theorem 1, $\widehat{\mathrm{CK}}\hat{R}^{\mathrm{es}} \subseteq \hat{I}$, which completes the proof. ☐

Proof of Theorem 4. Let $G = \mathrm{CB}R_i^{\mathrm{dm}}$. By Proposition 3 there exists an event $\bar{\Omega}$ which is the world perceived by all players. Let $F = G \cap \bar{\Omega}$. By Monderer and Samet (1989), Dist, and the definition of $\bar{\Omega}$, for each i, $F \subseteq G \subseteq \mathrm{B}_iG \subseteq \mathrm{B}_i(G \cap \bar{\Omega}) = \mathrm{B}_i(F)$.

Thus we can apply Lemma 1 to F. By (10), for $E \subseteq F$, $\hat{\mathrm{B}}_iE = (\mathrm{B}_iE) \cap F \subseteq (\mathrm{B}_iE) \cap \bar{\Omega}$. Since $\mathrm{B}_i((\neg\mathrm{B}_iE) \cup E) = \Omega$, the minimality of $\bar{\Omega}$ implies that $\bar{\Omega} \subseteq (\neg\mathrm{B}_iE) \cup E$. Thus, $(\mathrm{B}_iE) \cap \bar{\Omega} \subseteq (\mathrm{B}_iE) \cap ((\neg\mathrm{B}_iE) \cup E) = (\mathrm{B}_iE) \cap E) \subseteq E$. We conclude that for each $E \subseteq F$, $\hat{\mathrm{B}}_iE \subseteq E$, or equivalently, $(\neg\hat{\mathrm{B}}_iE) \cup E = \hat{\Omega}$. Thus, $\hat{\mathrm{K}}_i((\neg\hat{\mathrm{B}}_iE) \cup E) = \hat{\Omega}$. This equality with axiom K yields $\hat{\mathrm{K}}_i\hat{\mathrm{B}}_iE \subseteq \hat{\mathrm{K}}_iE$. Since by axiom KB2, we have $\hat{\mathrm{B}}_iE \subseteq \hat{\mathrm{K}}_i\hat{\mathrm{B}}_iE$, we conclude that $\hat{\mathrm{B}}_iE \subseteq \hat{\mathrm{K}}_iE$. The converse inclusion follows from KB1, and thus, $\hat{\mathrm{B}}_i = \hat{\mathrm{K}}_i$.

The proof that (1) holds for $\hat{\mathrm{K}}_i$ is the same as in Theorem 3. Using (10) and (11) we conclude that

$$(\mathrm{B}_i \neg \Omega^v) \cap F = \hat{\mathrm{B}}_i(\neg \Omega^v \cap F) = \hat{\mathrm{B}}_i \neg \hat{\Omega}^v,$$

and

$$(\neg \mathrm{B}_i(\neg \Omega^v \cup [h_i^v(\boldsymbol{s}; t_i) > h_i^v(\boldsymbol{s})]) \cap F = \neg \hat{\mathrm{B}}_i(\neg \hat{\Omega}^v \cup [h_i^v(\hat{\boldsymbol{s}}; t_i) > h_i^v(\hat{\boldsymbol{s}})]).$$

Thus, $R_i^{\mathrm{dm}} \cap F = \hat{R}_i^{\mathrm{dm}}$. Substituting $\hat{\mathrm{K}}_i$ for $\hat{\mathrm{B}}_i$, yields, $R_i^{\mathrm{dm}} \cap F = \hat{R}_i^{\mathrm{em}}$. By applying (12) to G, $G \subseteq \mathrm{B}_i R_i^{\mathrm{dm}}$. Therefore, $F = (\mathrm{B}_i R_i^{\mathrm{dm}}) \cap F = \hat{\mathrm{B}}_i(R_i^{\mathrm{dm}} \cap F) = \hat{\mathrm{B}}_i(\hat{R}_i^{\mathrm{em}}) = \hat{\mathrm{K}}_i(\hat{R}_i^{\mathrm{em}})$. Hence, $F = \mathrm{CK}\hat{R}_i^{\mathrm{em}}$, and by Theorem 2, $F \subseteq \hat{I} = I \cap F$. Finally, $G \subseteq \mathrm{B}_i F \subseteq \mathrm{B}_i(I)$ for each i, and therefore $G \subseteq \mathrm{CB}I$. $\square$

Proof of Proposition 3. First, we show that for each i, $\cap_E(\neg \mathrm{B}_i E \cup E)$ is the world perceived by i. By axiom K and Claim 3, $\mathrm{B}_i \cap_E (\neg \mathrm{B}_i E \cup E) = \Omega$. For the minimality, suppose $\mathrm{B}_i F = \Omega$. Then, $\cap_E(\neg \mathrm{B}_i E \cup E) \subseteq \neg \mathrm{B}_i F \cup F = F$. Thus, $\Omega_i = \cap_E(\neg \mathrm{B}_i E \cup E)$. By axiom K, axiom (Con) holds if and only if for each i and j, $\mathrm{B}_j(\cap_E(\neg \mathrm{B}_i E \cup E) = \Omega$, that is, $\mathrm{B}_j(\Omega_i) = \Omega$, which is equivalent to $\Omega_j \subseteq \Omega_i$. It is easy to see that t_i is a type function for i, p is a common prior of the for the type functions, and this probabilistic belief structure induces the believe structure. In particular, at each ω, $t_i(\omega)$ is equivalent to $t_i(\omega)$. $\square$

Proof of Proposition 4. Let $\bar{\Omega}_i$ be the world perceived by i. Then, $\bar{\Omega} = \cup_\omega b_i(\omega)$. Indeed, $B_i(\cup_\omega b_i(\omega)) = \Omega$ and thus $\bar{\Omega} \subseteq \cup_\omega b_i(\omega)$. Conversely, if $\omega' \notin \cup_\omega b_i(\omega)$, then, $\mathrm{B}_i(\bar{\Omega} \setminus \{\omega'\}) = \Omega$ contrary to the minimality of $\bar{\Omega}$.

Suppose that beliefs are interpersonally consistent, and let $\bar{\Omega}$ be the world perceived by all players. Let p be a probability function on Ω such that $p(\omega) > 0$ for each $\omega \in \bar{\Omega}$, and $p(\Omega \setminus \bar{\Omega}) = 0$. Thus, for each ω and i, $p(b_i(\omega)) > 0$. Define for each i and ω, $t_i(\omega)(\cdot) = p(\cdot \mid b_i(\omega))$. The probabilistic belief structure $(t_i)_i$ induces the belief structure. This shows that (1) implies (2). Obviously, for every probabilistic-belief structure $(t_i)_i$ with a common prior, the types in each state are equivalent. Thus (2) implies (3). Finally suppose that the belief structure is induced by a probabilistic belief structure $(t_i)_i$ with equivalent type functions. Define $\bar{\Omega}$ to be the set of all the states ω such that for some i (and hence for all i) $t_i(\omega) > 0$. Thus, $\bar{\Omega} = \cup_\omega b_i(\omega)$, and therefore $\bar{\Omega}$ is the world perceived by all players. $\square$

References

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

Aumann, R. J. (1995), Backward induction and common knowledge of rationality, *Games and Economic Behavior*, 8(1), 6–19.

Aumann, R. J. (1998), On the centipede game, *Games and Economic Behavior*, 23(1), 97–105.

Battigalli, P., and G. Bonanno, (1999), Recent results on belief, knowledge and the epistemic foundations of game theory, *Research in Economics*, 53(2), 149–225.

Ben-Porath, E. (1997), Rationality, Nash equilibrium and backwards induction in perfect-information games, *Review of Economic Studies*, 64, 23–46.

Bonanno, G., and K. Nehring. (1999), Assessing the truth axiom under incomplete information, *Mathematical Social Sciences*, 36(1), 3–29.

Halpern, J. Y., D. Samet, and E. Segev. (2009), Defining knowledge in terms of belief: The modal logic perspective, *The Review of Symbolic Logic*, 2, 469–487.

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

Monderer, D., and D. Samet. (1989), Approximating common knowledge with common beliefs, *Games and Economic Behavior*, 1(2), 170–190 .

Samet, D. (1996), Hypothetical knowledge and games with perfect information, *Games and Economic Behavior*, 17(2), 230–251.

Samet, D. (2011), Payoff-relevant rationality in games with perfect information, a manuscript.