Games and Economic Behavior 79 (2013) 192–200

Contents lists available at SciVerse ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Common belief of rationality in games of perfect information

Dov Samet

Tel Aviv University, Faculty of Management, Tel Aviv, Israel

ARTICLE INFO

Article history:
Received 12 June 2012
Available online 13 February 2013

JEL classification:
C70
C72

Keywords:
Perfect information
Common belief
Rationality
Backward induction
Centipede game

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

Common knowledge and common belief of rationality. Aumann (1995) proved that in perfect information games common knowledge of *substantive* rationality implies backward induction. Common knowledge of *material* rationality was shown in Aumann (1998) to imply backward induction in the centipede game. The language and model used in these papers are *epistemic* rather than *doxastic*, that is, they are formulated in terms of *knowledge* and not *belief*. Obviously, common knowledge is an epistemic notion. But even the notion of rationality is defined in both papers in terms of knowledge. For a player to be considered substantively rational it is required that for *each* vertex of hers she *does not know* that she could increase her conditional payoff at the vertex by deviating from her strategy. To be considered materially rational it is enough that this holds only for vertices that are reached.

The assumption of common knowledge is very demanding, not only because of the ever-increasing hierarchy of knowledge about knowledge, but also because of its epistemic nature. Knowledge, by virtue of being necessarily true, relates the internal realm of the mind and the outside world in mysterious ways, and more so where knowledge of *knowledge* is concerned. The idea that one does not err about the minds of others is mind boggling. No wonder that many researchers have studied the consequences of common belief of rationality rather than those of common knowledge of rationality.[1] Aumann (1995) tried to derive backward induction using common belief but clarified that the approach of his 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. Again, Aumann (1998) emphasized that the result was proved for knowledge and not for belief. We show in

E-mail address: dovsamet@gmail.com.

[1] Dozens of articles deal with the characterization of backward induction, the definition of rationality, and the implications of common knowledge and common belief of rationality. Comparing these articles is a daunting task, as the models used by different scholars vary vastly. A comparison of the results presented here to this literature would be beyond the scope of this paper. For a comprehensive survey of models of belief and knowledge and their applications to game theory, see Battigalli and Bonanno (1999). See also the survey on works on backward induction in Perea (2007).

0899-8256/$ – see front matter
http://dx.doi.org/10.1016/j.geb.2013.01.008

Example 1 below, that indeed, the main theorems of both papers do not hold when the weaker assumption of common *belief* of rationality is made rather than that of common *knowledge* of rationality.

Rationality defined in terms of belief. The reason Aumann failed to derive his results for common belief is that he applied common *belief* to rationality defined in terms of *knowledge*. Here we reexamine these results in purely doxastic language and model. This means, of course, that we have at our disposal only common belief and not common knowledge. However, it also means that rationality should be defined in terms of belief and not in terms of knowledge. This is done simply by replacing knowledge by belief in the definitions of rationality in Aumann (1995) and (1998). Thus, we say that a player is substantively rational *doxastically* if for *each* vertex of hers she *does not believe* that she could increase her conditional payoff at the vertex by deviating from her strategy, and materially rational *doxastically* if this is required only for vertices that are reached. The rationality in Aumann (1995) and (1998) would be referred to as *epistemic* rationality. With this change we state and prove the purely doxastic analogue of Aumann (1995):

Common belief of doxastic substantive rationality implies backward induction.

We now turn to 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 doxastic model we use consists of a state space with a belief operator for each player. We describe such models and characterize them axiomatically in terms of the properties of the belief operators. Such models can arise from partition models of knowledge in which each state is associated with probability functions that describe the probabilistic beliefs of the players. The belief operators in this case are the operators of belief with probability 1.

Comparing various notions of rationality. In a belief space each belief operator can be associated with a unique knowledge operator such that knowledge implies belief, and beliefs are known. Thus, we can compare doxastic and epistemic notions of rationality in the same model. Moreover, we can endow the belief space with probabilistic beliefs as explained above. With the probabilistic structure we can also define rationality as maximization of expected payoff, which we will refer to as *rationality by expectation*. The comparison of rationality by expectation with doxastic and epistemic substantive rationality in this model is straightforward. When a player maximizes her expected conditional payoff at a vertex of hers, then she cannot possibly believe with probability 1 that she can increase her conditional payoff by deviating from her strategy. And if she cannot believe this, then obviously she cannot know it. Thus, rationality by expectation implies doxastic substantive rationality, which implies epistemic substantive rationality. Therefore, the result of Aumann (1995) and its doxastic analogue here each imply that common belief of rationality by expectation implies backward induction. This observation was made in Aumann (1995) for epistemic substantive rationality. It explains why the epistemic model in Aumann (1995) and the doxastic model here do not make any use of probabilistic beliefs. Note, that the epistemic theorem of Aumann (1995) and the analogous doxastic theorem here are incomparable, as common belief is a weaker condition than common knowledge, but doxastic substantive rationality is stronger than epistemic rationality. Unlike the case of substantive rationality, the doxastic and epistemic versions of material rationality are not comparable as we explain later.

Conclusions. The results here show that in the model used in Aumann (1995) and (1998) knowledge and common knowledge are not crucial for the derivation of backward induction. In the model of Samet (1996), which formally introduces counterfactual strategic thinking, common knowledge of rationality is not sufficient to obtain backward induction. It seems then, that the variance in conclusions reached in different models of perfect information games is not related to the use of knowledge or belief. In accordance with previous findings, at least as far as game theoretic analysis is concerned, belief approximates knowledge and with the appropriate care can substitute for it.[2]

2. Common knowledge of epistemic rationality

We use the setup of Aumann (1995) and (1998). We consider a game of perfect information in general position. For such a game the result of backward induction is uniquely defined. The set of player i's vertices is denoted by V_i, and the set of i's strategies is S_i. The set of strategy profiles is denoted by S. For a vertex v, $h_i^v(s)$ is i's conditional payoff at v for $s \in S$.

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 $\mathrm{K}_i E = \{\omega \mid \Pi_i(\omega) \subseteq E\}$, where

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

$\Pi_i(\omega)$ is the element of Π_i that contains ω. The event that *all know* E is $\mathrm{K}E = \bigcap_i \mathrm{K}_i E$. 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, $\mathrm{CK}E = \bigcap_{n\geqslant 1} \mathrm{K}^n E$, where K^n is the nth power of K. Models of belief are described in the sequel.

A function $\boldsymbol{s}:\Omega \to S$ describes the strategy profile at each state. Player i's strategy function $\boldsymbol{s}_i$ is measurable with respect to her partition Π_i, which means that in each state each player knows his strategy. Events defined by a condition are described by enclosing the condition in square brackets. Thus, $[\boldsymbol{s}_i = s_i]$ consists of all states ω for which $\boldsymbol{s}_i(\omega) = s_i$. The event $[h_i^v(\boldsymbol{s}; t_i) > h_i^v(\boldsymbol{s})]$ consists of all states ω for which $h_i^v(\boldsymbol{s}(\omega); t_i) > h_i^v(\boldsymbol{s}(\omega))$ where $(\boldsymbol{s}(\omega); t_i)$ is the strategy profile obtained by changing i's strategy $\boldsymbol{s}_i(\omega)$ to t_i.

A standard assumption is made in this setup, that each player knows her strategy.

Knowing one's actual strategy. Each player knows she is playing 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 (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 at any of her vertices. Thus the event that i is substantively rational epistemically is:

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

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

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_i 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.

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

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

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

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

Theorem 2. *(See 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]

[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] Aumann (1995) *proves* that common knowledge of substantive rationality implies the backward-induction *strategies* but *states* the weaker claim that it implies the backward-induction *path*.

[5] Our presentation avoids the modal logic apparatus and uses set theoretic terminology instead. For a comprehensive discussion of these models from a modal logic perspective, see Battigalli and Bonanno (1999).

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

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, $b_i : \Omega \to 2^\Omega \setminus \{\emptyset\}$, 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 $B_i E = \{\omega \mid b_i(\omega) \subseteq E\}$. We think of $b_i(\omega)$ as the set of states that are considered possible by i at ω. By the measurability assumption, this set of states is considered possible in all the states in $\Pi_i(\omega)$. At each state in $\Pi_i(\omega) \setminus b_i(\omega)$, i is wrong thinking that the state lies in $b_i(\omega)$.

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* $(B_i^1)_i$ *be the operators associated with a probabilistic belief structure on* Ω*. The operators* $(B_i)_i$ *that are associated with the induced belief structure, satisfy* $B_i = B_i^1$.

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

(B1) $B(E \cap F) = BE \cap BF$ (distribution),
(B2) $BE \subseteq \neg B \neg E$ (consistency),
(B3) $BE \subseteq BBE$ (positive introspection),
(B4) $\neg BE \subseteq B \neg BE$ (negative introspection).

The distribution of belief over conjunction, as well as positive and negative introspection are well-known axioms of epistemic logic. Axiom B2 requires that belief is consistent in the sense that it is impossible to believe an event and its negation.

Note, that applying axiom B1 to events $E \subseteq F$ results in $BE = BE \cap BF$ which implies $BE \subseteq BF$. Thus, the operator B is monotonic.

These axioms are not only satisfied by the belief operators in belief structures, but also characterize them as we state next.

Proposition 1. *A set of operators* $B_i : 2^\Omega \to 2^\Omega$ *satisfy the axioms* B1–B4 *if and only if there exists a belief structure on* Ω *such that the belief operators associated with it are the operators* B_i.

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]:

(K1) $KE \subseteq E$ (truth)

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

(KB1) $K_i E \subseteq B_i E$
(KB2) $B_i E \subseteq K B_i E$

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 BE the event that all believe E, that is, $BE = \bigcap_i B_i E$, and by B^n, B to the power of n. The common belief operator is defined by $CBE = \bigcap_{n \geqslant 1} B^n E$. Since $K_i E \subseteq B_i E$, and as K_i and B_i are monotonic operators it follows that $CKE \subseteq CBE$. Thus, common belief is a weaker condition than common knowledge.

[7] In modal logic, axioms B1, B2, and B3 are usually denoted by D, 4, and 5, correspondingly. Distribution in modal logics is usually required over material implication, rather than conjunction, and is denoted by K. Thus, the logic of belief is referred to as the KD45 logic. Axiom B1 implies axiom K. In the logic KD45, axiom K implies B1 by virtue of the generalization inference rule. As we see, DK45 is also the logic of certainty, or belief in probability 1.

[8] By adding the truth axiom we can omit the axioms of contradiction and positive introspection which are derived from the axioms of distribution, negative introspection and truth.

[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. Common belief of doxastic rationality

Theorems 1 and 2, where rationality is defined in terms of knowledge, cannot be strengthened by replacing common knowledge of rationality by common knowledge of belief, 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. Thus, we first need to replace the operator K_i with B_i in (1). The assumption of knowing one's strategy, in (1), becomes,

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

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

This seems to be a relaxation of (1), since $E \subseteq \mathrm{B}_i E$ does not imply $E \subseteq \mathrm{K}_i E$ where K_i is the unique knowledge operator associated with B_i. However, since (4) is required for *all* strategies s_i, it follows that (4) does not relax (1), as we state next.

Proposition 2. *Condition* (4) *of* believing *one's actual strategy is equivalent to condition* (1) *of* knowing *one's actual strategy.*

The doxastic analogue of (2) is as follows:

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

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

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

By axiom KB1, doxastic substantive rationality is stronger than its 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,* $\mathrm{CB}R^{\mathrm{ds}} \subseteq I$.

The doxastic analogue of (3) is as follows:

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

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

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

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, as defined below, we can state a weaker version of this theorem.

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

4.1. Interpersonal consistency

The truth axiom states that knowledge is correct in the sense that if E is known it must hold. It can also be written equivalently as $\neg \mathrm{K}E \cup E = \Omega$, which says that the event "if E is known it holds" is necessarily true. For a belief operator B, $\neg \mathrm{B}E \cup E$ is not necessarily true (i.e., it does not have to be true in all states). However, $\neg \mathrm{B}E \cup E$ is necessarily *believed*.

Claim 2. *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 negative introspection and monotonicity $\neg \mathrm{B}(E) \subseteq \mathrm{B}(\neg \mathrm{B}(E)) \subseteq \mathrm{B}(\neg \mathrm{B}(E) \cup E)$. Thus, $\Omega \subseteq \mathrm{B}(\neg \mathrm{B}(E) \cup E)$.

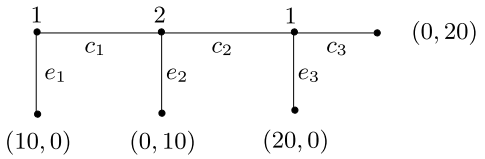

Fig. 1. A three-legged centipede game.

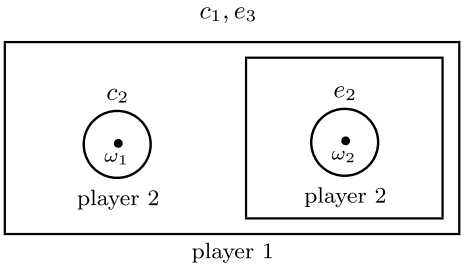

Fig. 2. Theorems 1 and 2 fail for common belief.

Claim 2 states that each player necessarily believes that her own beliefs are correct. We say that beliefs are interpersonally consistent when each player necessarily believes that like her, all other players have correct beliefs. Formally, beliefs in a belief structure are *interpersonally consistent* if for each i, j and E[11]:

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

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 of contradiction and distribution, $\mathcal{F} \neq \emptyset$ as $\neg \mathrm{B}_i \Omega \subseteq \mathrm{B}_i \neg \Omega = \emptyset$. By distribution, $\bigcap_{F \in \mathcal{F}} F \in \mathcal{F}$ and again by the axioms of contradiction and distribution 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.*
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 Fig. 1.

A model for this game is depicted in Fig. 2. The model describes the knowledge and belief of the the two players. Player 2's knowledge and belief coincide and they are described by the partition given by the circles. This player knows and believes in each state that this is indeed the state. Player 1's belief and knowledged are described by the boxes. The outer box describes player 1's knowledge, and both boxes describe player 1's belief. This player believes in both states to be in the second state, that is, $\mathrm{B}_1\{\omega_2\} = \Omega$.[13] The players' strategies are written above each element of the partitions.

[11] Bonanno and Nehring (1998) introduced this property under the more descriptive term "belief of no error". We have adopted the term consistency because it hints at 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.

[13] See Section 3.2 for the discussion and description of models of belief and knowledge.

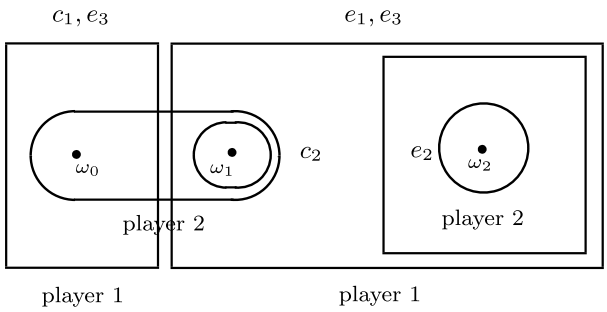

Fig. 3. The doxastic version of Theorem 2 fails.

- 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. This example shows that a straightforward doxastic version of Theorem 2, the main result of Aumann (1998), does not hold. Fig. 3 depicts a model of the game in Fig. 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 players 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 doxastically 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.

6. Proofs

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 $\beta(\omega) = \{E \mid \omega \in \text{B}E\}$ 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 $\beta(\omega) = \beta(\omega')$.

We prove that the partition Π can be described by,

$$\Pi(\omega) = \bigcap_{E \in \beta(\omega)} \text{B}E. \tag{7}$$

Obviously, for each BE such that $\omega \in \text{B}E$ it also holds that $\Pi(\omega) \subseteq \text{B}E$, which shows that $\Pi(\omega) \subseteq \bigcap_{E \in \beta(\omega)} \text{B}E$. Conversely, let $\omega' \in \bigcap_{E \in \beta(\omega)} \text{B}E$, then $\beta(\omega) \subseteq \beta(\omega')$. To show that equality in (7) holds, suppose to the contrary that for some E, $\omega' \in \text{B}E$ but $\omega \notin \text{B}E$. Then $\omega \in \neg\text{B}E$, and by negative introspection, $\omega \in \text{B}\neg\text{B}E$. Thus, $\neg\text{B}E \in \beta(\omega)$ which implies that $\omega' \in \text{B}\neg\text{B}E$. On the other hand, by positive introspection, $\omega' \in \text{BB}E$. This is impossible by the contradiction axiom. The equality $\beta(\omega) = \beta(\omega')$ implies that $\omega' \in \Pi(\omega)$, which completes the proof of (7).

Define for each ω, $b(\omega) = \bigcap_{E \in \beta(\omega)} E$. We show that for each ω and E, $\omega \in \text{B}E$ if and only if $b(\omega) \subseteq E$. For this we note that by distributivity and (7),

$$\text{B}b(\omega) = \bigcap_{E \in \beta(\omega)} \text{B}E = \Pi(\omega). \tag{8}$$

Suppose $b(\omega) \subseteq E$. By monotonicity $Bb(\omega) \subseteq BE$. By (8), $\Pi(\omega) \subseteq BE$ and thus, $\omega \in BE$. The converse implication follows immediately from the definition of b.

Finally, we show that (Π, b) is a belief structure. By definition b is measurable with respect to Π. To see that $b(\omega) \neq \emptyset$, note that ω is in the right-hand side of (8) and therefore $Bb(\omega) \neq \emptyset$. But this implies that $b(\omega) \neq \emptyset$, since by the axioms of distribution and contradiction, $B\emptyset = B(E \cap \neg E) = BE \cap B\neg E = \emptyset$. By distributivity and positive introspection, it follow from (7) that $B\Pi(\omega) = \Pi(\omega)$. Thus, $\omega \in B\Pi(\omega)$, which implies, as we have shown, that $b(\omega) \subseteq \Pi(\omega)$.

We have shown that b satisfies the three properties required to make (Π, b) a belief structure, and that B is the operator associated with it. □

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

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

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

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

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

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

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

Proof of Proposition 2. By axiom KB1, for each s_i, $K_i[\boldsymbol{s}_i = s_i] \subseteq B_i[\boldsymbol{s}_i = s_i]$. For the converse inclusion, observe that for $s_i \neq s_i'$, $[\boldsymbol{s}_i = s_i] \cap [\boldsymbol{s}_i = s_i'] = \emptyset$. Since we have already shown that $B_i(\emptyset) = \emptyset$ it follows by distribution that $B_i[\boldsymbol{s}_i = s_i] \cap B_i[\boldsymbol{s}_i = s_i'] = \emptyset$. But $([\boldsymbol{s}_i = s_i])_{\bigcup_{s_i \in S_i}}$ is a partition of Ω. Thus (4) implies that for each s_i, $[\boldsymbol{s}_i = s_i] = B_i[\boldsymbol{s}_i = s_i]$. By axiom KB2, $B_i[\boldsymbol{s}_i = s_i] \subseteq K_i B_i[\boldsymbol{s}_i = s_i]$. Substituting in this inclusion $B_i[\boldsymbol{s}_i = s_i]$ for the equal terms $[\boldsymbol{s}_i = s_i]$ we get the desired inclusion. □

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

$$F \subseteq B_i F \cap B_i R_i^{ds}. \tag{12}$$

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 $\hat{K}_i$.

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

By (11) and axiom KB1,

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

Thus, for each i,

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

where the last event with the hat is i's rationality as defined in (2) for $\hat{K}_i$. Since K_i satisfies the axioms of distribution, negative introspection and truth,

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

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

$$BKE \subseteq KE, \tag{16}$$

because by negative introspection and KB1, $BKE \cap \neg KE = BKE \cap K\neg KE \subseteq BKE \cap B\neg KE$, which by the axioms of distribution and contradiction is empty.

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

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

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

and

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

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. By Monderer and Samet (1989), G, being the common belief of an event, is an evident belief, and the inclusions $G \subseteq \mathrm{B}_i(I)$ imply that $G \subseteq \mathrm{CB}I$. □

Proof of Proposition 3. Let Ω_i be the world perceived by i. First, we show that for each i, $\Omega_i = \bigcap_E (\neg \mathrm{B}_i E \cup E)$. By axiom Dist and Claim 2, $\mathrm{B}_i(\bigcap_E (\neg \mathrm{B}_i E \cup E)) = \Omega$. For the minimality, suppose $\mathrm{B}_i F = \Omega$. Then, $\bigcap_E (\neg \mathrm{B}_i E \cup E) \subseteq \neg \mathrm{B}_i F \cup F = F$.

By distributivity, axiom Con holds if and only if for each i and j, $\mathrm{B}_j(\bigcap_E (\neg \mathrm{B}_i E \cup E)) = \Omega$, that is, $\mathrm{B}_j(\Omega_i) = \Omega$, which is equivalent to $\Omega_j \subseteq \Omega_i$. □

Proof of Proposition 4. Let $\bar{\Omega}_i$ be the world perceived by i. Then, $\bar{\Omega}_i = \bigcup_\omega b_i(\omega)$. Indeed, $B_i(\bigcup_\omega b_i(\omega)) = \Omega$ and thus $\bar{\Omega}_i \subseteq \bigcup_\omega b_i(\omega)$. Conversely, since $B_i(\bar{\Omega}_i) = \Omega$, it follows that for each ω, $b_i(\omega) \subseteq \bar{\Omega}_i$.

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, $\bar{\Omega} = \bigcup_\omega b_i(\omega)$, and therefore $\bar{\Omega}$ is the world perceived by all players. □

References

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

Aumann, R.J., 1998. On the centipede game. Games Econ. Behav. 23 (1), 97–105.

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

Bonanno, G., Nehring, K., 1998. Assessing the truth axiom under incomplete information. Math. Soc. Sci. 36 (1), 3–29.

Halpern, J.Y., Samet, D., Segev, E., 2009. Defining knowledge in terms of belief: The modal logic perspective. Rev. Symbol. Log. 2, 469–487.

Monderer, D., Samet, D., 1989. Approximating common knowledge with common beliefs. Games Econ. Behav. 1 (2), 170–190.

Perea, A., 2007. Epistemic foundations for backward induction: An overview. In: Van Bentham, J., Gabbay, D., Löwe, B. (Eds.), Interactive Logic, Selected Papers from the 7th Augustus de Morgan Workshop, London. Amsterdam University Press.

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

Samet, D., 2011. On the dispensable role of time in games of perfect information. A manuscript.