Synthese (2007) 158:251–271
Knowledge, Rationality & Action 341–361
DOI 10.1007/s11229-007-9217-2

A one-person doxastic characterization of Nash strategies

Andrés Perea

Received: 22 February 2007 / Accepted: 26 June 2007 / Published online: 21 July 2007

Abstract Within a formal epistemic model for simultaneous-move games, we present the following conditions: (1) *belief in the opponents' rationality* (BOR), stating that a player believes that every opponent chooses an optimal strategy, (2) *self-referential beliefs* (SRB), stating that a player believes that his opponents hold correct beliefs about his own beliefs, (3) *projective beliefs* (PB), stating that i believes that j's belief about k's choice is the same as i's belief about k's choice, and (4) *conditionally independent beliefs* (CIB), stating that a player believes that opponents' types choose their strategies independently. We show that, if a player satisfies BOR, SRB and CIB, and believes that every opponent satisfies BOR, SRB, PB and CIB, then he will choose a Nash strategy (that is, a strategy that is optimal in some Nash equilibrium). We thus provide a sufficient collection of one-person conditions for Nash strategy choice. We also show that none of these seven conditions can be dropped.

Keywords Nash equilibrium · Epistemic game theory

1 Introduction

Since its introduction by Nash (1951), the concept of Nash equilibrium has played an essential role in game theory and its various applications. It is therefore natural to look for reasonable conditions under which players may be expected to choose according to Nash equilibrium. Such conditions have been provided in many different settings. In learning theory one assumes that players play the game repeatedly, and one can find

A. Perea (✉)
Department of Quantitative Economics, Maastricht University,
P.O. Box 616, Maastricht, 6200 MD, The Netherlands
e-mail: a.perea@ke.unimaas.nl
http://www.personeel.unimaas.nl/a.perea/

reasonable classes of learning rules that eventually lead players to choose according to some Nash equilibrium. Similarly, evolutionary game theory studies classes of replicator dynamics in the repeated game that converge to (special types of) Nash equilibria in the long run. If the game is to be played only once, one could think of a situation in which a mediator publicly announces the mixed strategy profile to the players, and recommends each player to play his strategy in this profile. Then, players may only be expected to follow their recommendation if the mixed strategy profile is a Nash equilibrium. One can also formulate sufficient conditions for Nash equilibrium in a static setting without mediators. See, for instance, Brandenburger and Dekel (1987), Aumann and Brandenburger (1995) and Asheim (2006). A key condition in each of these models is that a player knows, or has a correct belief about, his opponents' beliefs about the other players' strategy choices.

A common feature of each of the models above is that it requires, either explicitly or implicitly, some sort of communication between players. In the learning models and evolutionary models players communicate by the actions they choose in the repeated game. It is this type of communication that allows them to converge to a Nash equilibrium in the long run. In the static models where players are assumed to have correct beliefs about the other players' beliefs, there seems to be a need for ex-ante communication between players in which they report their beliefs to others. Otherwise, there is no reason to expect that players should be right about the opponents' beliefs, even if common belief in rationality is imposed.

In this paper, we completely focus on static settings in which there is no communication between players. In such a setting, a player can base his strategy choice solely upon his own beliefs about the opponents' choices and his own beliefs about the opponents' beliefs, since before making his decision he receives no information about what opponents do or believe. It thus makes sense to analyze the game from a single player's perspective, and this is exactly what we will do. More precisely, we focus on one player, say player i, in the game, and impose conditions solely on the beliefs that player i has about the other players' strategy choices and the other players' beliefs. We do not impose any conditions on the beliefs and choices of players other than i. We thus allow for the event that player i believes that some opponent believes or chooses in a certain way, whereas in fact this opponent believes or chooses differently.

A question that arises within this one-person approach is: "How should we interpret Nash equilibrium?" Usually, Nash equilibrium is either interpreted as an equilibrium in choices or as an equilibrium in beliefs. The first interpretation states that a player's choice should be optimal given the other players' choices, whereas the second interpretation states that player i's belief about player j's choice should only assign positive probability to choices that are optimal for player j, given j's belief about the other players' choices. Both interpretations, however, require that we simultaneously impose conditions on the choices or beliefs of *all* players, not just player i. So, how can we make sense of Nash equilibrium in a one-person approach?

A possible way is to interpret all the beliefs in a Nash equilibrium as belonging to a single player. Consider, for instance, a game with two players, say players i and j, and a Nash equilibrium (μ_i, μ_j), where μ_i is a probability distribution over i's choices, and μ_j is a probability distribution over j's choices. A possible one-person interpretation is that μ_j represents i's belief about j's choice, and that μ_i represents i's belief about j's belief about i's choice. That is, the Nash equilibrium (μ_i, μ_j) is completely "in the mind of player i," as μ_i and μ_j represent beliefs that both belong to him. The Nash equilibrium, when interpreted in this way, says nothing about player j's actual belief about i's choice. In fact, it may be the case that in this Nash equilibrium, player i is completely wrong about what j really believes or does.

Once we interpret Nash equilibrium as describing player i's state of mind, we may explore its behavioral consequences. So, again in a two player setting, if i's belief about j's choice and i's belief about j's belief about i's choice together constitute a Nash equilibrium, what can i rationally choose? The answer is almost tautological: any strategy choice for i that is optimal in some Nash equilibrium (μ_i, μ_j). We refer to such strategies as *Nash strategies*. The logical relation between Nash strategies on the one hand, and strategies receiving positive probability in some Nash equilibrium on the other hand is more subtle than it may appear at first. Every strategy that receives positive probability in some Nash equilibrium is a Nash strategy, but not every Nash strategy receives positive probability in some Nash equilibrium. Consider, for instance, the game in Fig. 1. Player 1 is the row player, and player 2 the column player. In this game, $(b, \frac{1}{2}c + \frac{1}{2}d)$ is a Nash equilibrium. Since strategy a is optimal against $\frac{1}{2}c + \frac{1}{2}d$, it follows that a is a Nash strategy. However, there is no Nash equilibrium (μ_1, μ_2) in which μ_1 assigns positive probability to a. At the same time, we should not exclude the *choice a* if we require that player 1's belief about player 2's choice and player 1's belief about player 2's belief about his own choice constitute a Nash equilibrium. It is possible, namely, that player 1's belief about 2's choice is given by $\frac{1}{2}c + \frac{1}{2}d$, that he believes that 2 believes that he will choose b, whereas in fact he will choose a. Then, player 1's beliefs above constitute a Nash equilibrium, but he chooses a Nash strategy that cannot have positive probability in any Nash equilibrium.

The objective of this paper is to identify reasonable conditions on the beliefs of a single player which would guarantee that this player eventually chooses a Nash strategy. We thus offer one-person doxastic conditions for *Nash strategies*. As an illustration of our objective, consider the game in Fig. 2, which is taken from Myerson (1991,

	c	d
a	2, 0	0, 2
b	1, 1	1, 1

Fig. 1 Not every Nash strategy is assigned positive probability in a Nash equilibrium

	d	e	f
a	3, 0	0, 2	0, 3
b	2, 0	1, 1	2, 0
c	0, 3	0, 2	3, 0

Fig. 2 Rationalizability versus Nash strategies

p. 94). Suppose that the game is played only once, and that player 1 and player 2 cannot communicate to each other. We will analyze this game completely from player 1's perspective. It can be shown that common belief in rationality does not exclude any of player 1's strategies, but that strategy b is the only Nash strategy for player 1. Consider, namely, a scenario in which

1. player 1 believes that player 2 chooses d,
2. player 1 believes that player 2 believes that player 1 chooses c,
3. player 1 believes that player 2 believes that player 1 believes that player 2 chooses f,
4. player 1 believes that player 2 believes that player 1 believes that player 2 believes that player 1 chooses a,
5. player 1 believes that player 2 believes that player 1 believes that player 2 believes that player 1 believes that player 2 chooses d,

and so on. Then, player 1 respects common belief in rationality, since he believes that player 2 chooses rationally, believes that player 2 believes that player 1 chooses rationally, and so on. Since in this scenario it is rational for player 1 to choose a, we must allow for the strategy a if we only impose common belief in rationality. Since strategies a and c play similar roles in this game, one can construct a similar scenario that leads to strategy choice c. Finally, strategy b can be justified by a much simpler scenario in which

1. player 1 believes that player 2 chooses e,
2. player 1 believes that player 2 believes that player 1 chooses b,
3. player 1 believes that player 2 believes that player 1 believes that player 2 chooses e,

and so on. Summarizing, every strategy for player 1 can be justified by a belief hierarchy for player 1 that respects common belief in rationality. More generally, Tan and Werlang (1988) have shown that the strategies which can be chosen rationally in two-player games when imposing only common belief in rationality are exactly the *rationalizable strategies* in the sense of Bernheim (1984) and Pearce (1984).[1]

On the other hand, it can be shown that b is the only Nash strategy for player 1. In order to see this, it is helpful to consider Fig. 3 which depicts the players' best response correspondences in this game. The first triangle should be read as follows. Every point in the triangle represents a probabilistic belief of player 1 about player 2's strategies. The three areas in this triangle represent the sets of beliefs for which the strategies a, b and c are optimal, respectively. Similarly for the second triangle. Consider a Nash equilibrium (μ_1, μ_2) of the game, where μ_i is a probability distribution over i's strategies for $i = 1, 2$. We show that μ_1 must assign probability 1 to b, and μ_2 must assign probability 1 to e. Suppose, contrary to what we want to show, that μ_1 assigns positive probability to a. Then, since (μ_1, μ_2) is a Nash equilibrium, a must be optimal for player 1 against μ_2. By Fig. 2, μ_2 must then assign positive probability to d, and hence d must be optimal against μ_1. By Fig. 2, μ_1 must then

[1] For more than two players, one needs to impose that beliefs about different opponents be stochastically independent in order to obtain the equivalence with rationalizability.

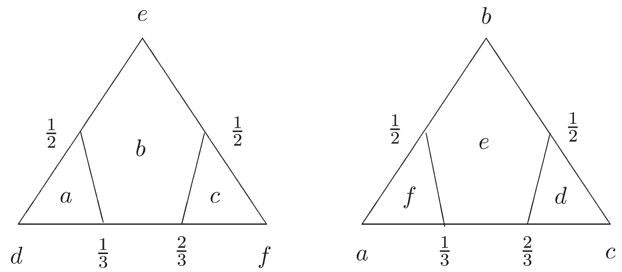

Fig. 3 Best response correspondences for the game in Fig. 2

assign positive probability to c, and hence c must be optimal against μ_2. So, both a and c should be optimal against μ_2. However, from the first triangle it is clear that a and c cannot both be optimal against μ_2, and hence we have a contradiction. By symmetry, one can similarly prove that μ_1 cannot assign positive probability to c, and that μ_2 cannot assign positive probability to d or f. Hence, the only strategy that can rationally be chosen by player 1 in a Nash equilibrium is b, and therefore b is the only Nash strategy for player 1.

Now, let us compare the scenario above which led to the non-Nash strategy a with the simpler scenario that led to the Nash strategy b. One fundamental difference between both scenarios is that the first assumes that player 1 believes that player 2 is wrong about 1's belief, whereas the second scenario assumes that player 1 believes that player 2 is right about 1's belief. Namely, in the first scenario player 1 believes that player 2 chooses d, but he believes that player 2 believes that he believes that player 2 chooses f, and not d.

Even stronger, the choice a can *only* be justified by scenarios in which player 1 believes that player 2 is wrong about 1's belief. In order to see this, assume that strategy a would be optimal for player 1. Then, player 1 must believe with positive probability that player 2 chooses d. In turn, player 2 can only rationally choose d if he believes with positive probability that player 1 chooses c. Hence, player 1 must believe with positive probability that player 2 believes with positive probability that player 1 chooses c. However, there is no belief for player 1 for which both a and c are rational, and hence player 1 must believe with positive probability that player 2 believes with positive probability that player 1's belief is different from his true belief.

Say that a belief hierarchy β_i for player i has *self-referential beliefs* if β_i believes that every opponent j believes that i's belief hierarchy is β_i. For the example in Fig. 2, we may thus conclude that the difference between Nash strategies and rationalizable non-Nash strategies is that the former can be justified by a self-referential belief hierarchy, whereas the latter cannot. The question is whether this is true generally. In Theorem 4.5 we show that this is true for the class of two-player games. More precisely, we show in Theorem 4.5 that in every two-player game, a player who (1) believes in the opponent's rationality (BOR), (2) has self-referential beliefs (SRB), (3) believes that his opponent BOR, and (4) believes that his opponent has SRB, must choose a Nash strategy.

The four conditions above are no longer enough if we turn to more than two players. Similarly to existing foundations for Nash equilibrium in the literature, we encounter the following two problems for more than two players: (a) We must guarantee that player i's belief about opponent j's strategy choice should be stochastically independent from i's belief about opponent k's strategy choice, and (b) we must guarantee that player i believes that opponents j and k hold the same belief about player l's strategy choice. In this paper, we guarantee these events by imposing *conditionally independent beliefs* and *projective beliefs*. Conditionally independent beliefs (CIB) is taken from Brandenburger and Friedenberg (2006), and says that for any given profile of belief hierarchies for i's opponents, player i's belief about his opponents' strategy choices *conditional on this profile of belief hierarchies* should be uncorrelated. In other words, any correlation in i's belief about his opponents' strategy choices should be due to correlation in his belief about the opponents' belief hierarchies. The idea behind this condition is that players with fixed belief hierarchies are assumed to choose their strategies independently. Projective beliefs (PB), in turn, states that player i believes that j's belief about player k is the same as his own belief about player k. That is, player i projects his own belief about player k on player j. In Theorem 4.5 we show that for three players or more, a player who (1) BOR, (2) has SRB, (3) believes that his opponents BOR, (4) believes that his opponents have SRB, and, in addition, (5) has CIB, (6) believes that his opponents have PB, and (7) believes that his opponents have CIB, must choose a Nash strategy. At this stage, one may wonder why we did not impose that the player himself has PB. The reason is that this property follows from the assumptions that the player has SRB and believes that every opponent has PB (see Lemma 4.2).

The outline of this paper is as follows: In Sect. 2 we present our epistemic model. Section 3 formally introduces the notions of BOR, SRB, PB and CIB. In Sect. 4 we show that every player who satisfies the conditions (1)–(7) above must choose a Nash strategy. In Theorem 4.7 we prove the converse of this result, namely that every Nash strategy can rationally be chosen by a player who satisfies these conditions (1)–(7). In Sect. 5 we prove that none of these seven conditions can be dropped in Theorem 4.5. In particular, this implies that the seven conditions are logically independent. In Sect. 6 we discuss the intuitive content of Theorems 4.5 and 4.7, with a special focus on SRB and PB. In Sect. 7 we compare our conditions with sufficient conditions for Nash equilibrium as provided in Aumann and Brandenburger (1995), Asheim (2006) and Brandenburger and Dekel (1987).

2 Epistemic model

Let I be a finite set of players. A finite *game* is a tuple $\Gamma = (S_i, u_i)_{i\in I}$, where S_i is the finite set of *strategies* for player i, and $u_i\colon \times_{j\in I} S_j \to \mathbb{R}$ is player i's *utility function*. We shall assume throughout that a player believes that the utility functions are as specified by Γ, that a player believes that all players believe this, and so on. For every finite set X, let $\Delta(X)$ denote the set of probability distributions on X.

Definition 2.1 (Epistemic model) A finite *epistemic model* for the game Γ is a tuple

$$\mathcal{M} = (T_i, \beta_i)_{i\in I}$$

where, for every player i, T_i is a finite set of types, and β_i is a one-to-one function from T_i to $\Delta(S_{-i} \times T_{-i})$.

Here, S_{-i} is a short way to write $\times_{j \neq i} S_j$, and similarly for T_{-i}. The interpretation of β_i is that for every type $t_i \in T_i$, the image $\beta_i(t_i)$ denotes t_i's probabilistic belief about the opponents' strategy choices and types. For every event $E \subseteq S_{-i} \times T_{-i}$ for player i and every number $p \in [0, 1]$, we say that type t_i *believes* the event E *with probability* p if $\beta_i(t_i)(E) = p$. We say that t_i *believes* E if $\beta_i(t_i)(E) = 1$. For instance, we say that t_i believes that player j has type t_j if t_i assigns probability 1 to the set of strategy-type profiles in $S_{-i} \times T_{-i}$ where player j's type is t_j.

3 Restrictions on beliefs

In this section we discuss four conditions that one may impose on a player's beliefs: *belief in the opponents' rationality, self-referential beliefs, projective beliefs* and *conditionally independent beliefs*. We first need some additional terminology. For every type t_i and strategy profile $s_{-i} \in S_{-i}$, we denote by $\beta_i(t_i)(s_{-i})$ the probability that the belief $\beta_i(t_i)$ assigns to the set $\{s_{-i}\} \times T_{-i}$. For every strategy $s_i \in S_i$, we denote by

$$u_i(s_i, t_i) := \sum_{s_{-i} \in S_{-i}} \beta_i(t_i)(s_{-i}) u_i(s_i, s_{-i})$$

the expected utility for type t_i of choosing strategy s_i. We say that strategy s_i is *rational* for type t_i if $u_i(s_i, t_i) \geq u_i(s_i', t_i)$ for every $s_i' \in S_i$.

Definition 3.1 (Belief in the opponents' rationality) Type t_i is said to *believe in the opponents' rationality* if for every opponent j, and every strategy-type pair $(s_j, t_j) \in S_j \times T_j$ to which t_i assigns positive probability, the strategy s_j is rational for type t_j.

Definition 3.2 (Self-referential beliefs) Type t_i is said to have *self-referential beliefs* if for every $p \in [0, 1]$ and every event $E \subseteq S_{-i} \times T_{-i}$ which t_i believes with probability p, type t_i believes that all opponents believe that player i believes E with probability p.

Hence, a player with self-referential beliefs thinks that his opponents hold correct beliefs about his own beliefs. Now, let j be an opponent of i, and let $E_j \subseteq S_j \times T_j$ be an event about player j. We say that t_i believes E_j with probability p if t_i assigns probability p to the event

$$E_j \times (\times_{k \neq i,j}(S_k \times T_k)).$$

Definition 3.3 (Projective beliefs) Type t_i is said to have *projective beliefs* if for every pair of opponents j, k and every event $E_k \subseteq S_k \times T_k$ about player k: if t_i believes E_k with probability p, then t_i believes that j believes E_k with probability p.

Intuitively, a player with projective beliefs projects his belief about an opponent on his other opponents. Of course, this condition only imposes restrictions if there are at least three players. Our last condition, *conditionally independent beliefs*, is taken from Brandenburger and Friedenberg (2006). It states that for every given profile t_{-i} of opponents' types which is deemed possible by type t_i, his belief about the opponents' strategies *conditional on* t_{-i} should be uncorrelated. In other words, any correlation in t_i's belief about the opponents' strategy choices should come from correlation in his belief about the opponents' types. The idea behind this condition is that types are assumed to choose their strategies independently since pre-play communication is not allowed. However, player i's belief about j's choice may still be dependent on his belief about k's choice if his belief about j's type is dependent on his belief about k's type. To formalize this condition, we need some terminology. Let t_{-i} be a profile of opponents' types to which t_i assigns positive probability, and let s_{-i} be a profile of opponents' strategies. By

$$\beta_i(t_i)(s_{-i}|t_{-i}) := \frac{\beta_i(t_i)(s_{-i}, t_{-i})}{\beta_i(t_i)(t_{-i})}$$

we denote the probability that t_i assigns to the strategy profile s_{-i}, conditional on the event that the opponents' types are given by t_{-i}. Here, $\beta_i(t_i)(t_{-i})$ denotes the probability that t_i assigns to the event that the opponents' types are t_{-i}. Similarly, for every opponent j, every type t_j to which t_i assigns positive probability, and every strategy $s_j \in S_j$, we denote by

$$\beta_i(t_i)(s_j|t_j) := \frac{\beta_i(t_i)(s_j, t_j)}{\beta_i(t_i)(t_j)}$$

the probability that t_i assigns to strategy choice s_j, conditional on the event that j's type is t_j. Here, $\beta_i(t_i)(s_j, t_j)$ denotes the probability that t_i assigns to the event that j chooses s_j and has type t_j, whereas $\beta_i(t_i)(t_j)$ is the probability that t_i assigns to the event that j has type t_j.

Definition 3.4 (Conditionally independent beliefs) Type t_i is said to have *conditionally independent beliefs* if for every $t_{-i} \in T_{-i}$ with $\beta_i(t_i)(t_{-i}) > 0$, and every $s_{-i} \in S_{-i}$:

$$\beta_i(t_i)(s_{-i}|t_{-i}) = \prod_{j \neq i} \beta_i(t_i)(s_j|t_j).$$

In fact, this condition combines the restrictions of *conditional independence* and *sufficiency* in Brandenburger and Friedenberg (2006). In their model, sufficiency states that player i's belief about player j's strategy choice should be independent of his belief about some other opponent's type. Obviously, our notion of conditionally independent beliefs satisfies this additional requirement. It should be clear that conditionally independent beliefs only imposes restrictions if there are at least three players.

As an abbreviation, we denote by BOR, SRB, PB and CIB the events that types believe in the opponents' rationality, have self-referential beliefs, have projective beliefs, and have conditionally independent beliefs, respectively. We say that type t_i believes that every opponent satisfies BOR if $\beta_i(t_i)$ only assigns positive probability to opponents' types that satisfy BOR. Similarly, we define the events that t_i believes that every opponent satisfies SRB, PB and CIB.

4 Relation with Nash strategies

In this section we show that every type which satisfies BOR, SRB, and CIB, believes that every opponent satisfies BOR, SRB, PB and CIB, and chooses rationally, must choose a Nash strategy. A profile $(\mu_i)_{i\in I}$ of probability distributions $\mu_i \in \Delta(S_i)$ is called a *Nash equilibrium* for the game Γ if, for every player i, $\mu_i(s_i) > 0$ only if $u_i(s_i, \mu_{-i}) \geq u_i(s_i', \mu_{-i})$ for every $s_i' \in S_i$. Here,

$$u_i(s_i, \mu_{-i}) := \sum_{s_{-i}\in S_{-i}} \prod_{j\neq i} \mu_j(s_j) u_i(s_i, s_{-i})$$

denotes the expected utility for player i of choosing s_i if his belief about the opponents' strategies is given by μ_{-i}.

We call a strategy s_i a *Nash strateg*y for Γ if there is a Nash equilibrium $(\mu_i)_{i\in I}$ for Γ such that $u_i(s_i, \mu_{-i}) \geq u_i(s_i', \mu_{-i})$ for every $s_i' \in S_i$. Since this definition is non-standard, it requires some further explanation. In many models, a Nash equilibrium $(\mu_i)_{i\in I}$ is interpreted as an equilibrium in *beliefs*, and μ_{-i} represents player i's belief about his opponents' strategy choices. This is also the viewpoint I take in this paper. With this interpretation in mind, one can ask the following natural question: What are the behavioral consequences of requiring that player i's belief about the opponents' strategy choices is part of a Nash equilibrium in beliefs? In other words, which strategies can player i rationally choose if his belief μ_{-i} about the opponents' strategies is part of a Nash equilibrium in beliefs? The answer must then be: any Nash strategy as defined above. Hence, the definition of a Nash strategy formalizes the behavioral consequences of assuming that player i's beliefs about the opponents is part of a Nash equilibrium in beliefs. Recall from the introduction that every strategy that is assigned positive probability in some Nash equilibrium in beliefs is also a Nash strategy, but the converse is not true.

Before we prove our theorem on the relation with Nash strategies, we first derive some implications of SRB, PB, CIB, and belief in these events. In each of these lemmas, we assume that $\mathcal{M} = (T_i, \beta_i)_{i\in I}$ is a finite epistemic model for a finite game Γ.

Lemma 4.1 *Let $t_i \in T_i$ be a type with SRB. Then, t_i believes that every opponent believes that i's type is t_i.*

Proof Choose an arbitrary event $E \subseteq S_{-i} \times T_{-i}$ for player i, and assume that t_i believes E with probability p. By SRB, t_i believes that every opponent believes that i believes E with probability p. Since this holds for every E, and since the function $\beta_i\colon T_i \to \Delta(S_{-i} \times T_{-i})$ in the epistemic model is one-to-one, t_i believes that every opponent believes that i's type is t_i. □

Lemma 4.2 *Let $t_i \in T_i$ be a type with SRB which believes that every opponent has PB. Then, t_i has PB.*

Proof Suppose that j and k are two different opponents of player i. Let $E_k \subseteq S_k \times T_k$ be an event which t_i believes with probability p, and let $t_j \in T_j$ be a type to which t_i assigns positive probability. We show that t_j believes E_k with probability p as well.

Since t_i has SRB, we know from Lemma 4.1 that t_j believes that i's type is t_i. Since t_i believes that j has PB, it must be the case that t_j has PB, and hence t_j's belief about player k must be the same as t_i's belief about player k. Consequently, t_j must believe E_k with probability p. We may thus conclude that t_i has PB. □

Lemma 4.3 *Let t_i be a type with SRB which believes that every opponent has PB and SRB. Then, there are opponents' types $t_k \in T_k$, $k \in I\backslash\{i\}$, such that (1) t_i believes for every $j \neq i$ that j's type is t_j, (2) for every $j \neq i$, type t_j believes that i's type is t_i, and (3) for every $j, k \neq i$, type t_j believes that k's type is t_k.*

Proof Suppose that $j \neq i$, and that t_i assigns positive probability to type $t_j \in T_j$. By Lemma 4.1 we know that t_i believes that j believes that i's type is t_i. Hence, t_j must believe that i's type is t_i. Since t_i believes that j has SRB, type t_j must have SRB. By applying Lemma 4.1 to t_j, we may conclude that t_j believes that i believes that j's type is t_j. Since we have seen that t_j believes that i's type is t_i, it follows that t_i must believe that j's type is t_j. Hence, we have shown properties (1) and (2). It remains to show (3). By Lemma 4.2 we know that t_i has PB. Suppose that $j, k \neq i$. Since t_i believes that k's type is t_k, and since t_i has PB, it follows that t_i believes that j believes that k's type is t_k. Since t_i believes that j's type is t_j, type t_j believes that k's type is t_k. This completes the proof. □

Lemma 4.4 *Let $t_i \in T_i$ be a type that has SRB and CIB, and believes that every opponent has SRB, PB and CIB. Then, there are probability distributions $\mu_k \in \Delta(S_k)$, $k \in I$, such that (1) t_i believes that every opponents' strategy profile $(s_j)_{j\neq i}$ is chosen with probability $\prod_{j\neq i} \mu_j(s_j)$, (2) for every $j \neq i$, type t_i believes that j believes that every opponents' strategy profile $(s_k)_{k\neq j}$ is chosen with probability $\prod_{k\neq j} \mu_k(s_k)$, and (3) for every $j \neq i$, type t_i believes that j believes that i believes that every opponents' strategy profile $(s_j)_{j\neq i}$ is chosen with probability $\prod_{j\neq i} \mu_j(s_j)$.*

Proof By Lemma 4.3, there are opponents' types $t_k \in T_k$, $k \in I\backslash\{i\}$, such that (1) t_i believes for every $j \neq i$ that j's type is t_j, (2) for every $j \neq i$, type t_j believes that i's type is t_i, and (3) for every $j, k \neq i$, type t_j believes that k's type is t_k. Let $t_{-i} := (t_j)_{j\neq i}$. Since t_i has CIB, it holds that

$$\beta_i(t_i)(s_{-i}|t_{-i}) = \prod_{j\neq i} \beta_i(t_i)(s_j|t_j) \tag{4.1}$$

for every $s_{-i} \in S_{-i}$. Now, define $\mu_j(s_j) := \beta_i(t_i)(s_j|t_j)$ for every $j \neq i$ and every $s_j \in S_j$. Then, by (4.1) and the fact that t_i believes that every opponent j has type t_j, type t_i believes that every opponents' strategy profile s_{-i} is chosen with probability $\prod_{j\neq i} \mu_j(s_j)$. We have thus shown property (1) of this lemma.

Now, choose some fixed opponent $j \neq i$. Since t_i believes that j has CIB, and t_i believes that j's type is t_j, type t_j has CIB. Hence,

$$\beta_j(t_j)(s_{-j}|t_{-j}) = \prod_{k\neq j} \beta_j(t_j)(s_k|t_k) \tag{4.2}$$

for every opponents' strategy profile s_{-j}. Choose some arbitrary player $k \notin \{i, j\}$. Since, by Lemma 4.2, t_i has PB and believes that j's type is t_j, type t_i's belief about k's strategy choice must be equal to t_j's belief about k's strategy choice, and hence $\beta_j(t_j)(s_k|t_k) = \beta_i(t_i)(s_k|t_k) = \mu_k(s_k)$ for every $s_k \in S_k$. Define $\mu_i(s_i) := \beta_j(t_j)(s_i|t_i)$ for every $s_i \in S_i$. Together with (4.2), we may then conclude that t_j believes that every opponents' strategy profile s_{-j} is chosen with probability $\prod_{k \neq j} \mu_k(s_k)$. So, t_i believes that j believes that every opponents' strategy profile s_{-j} is chosen with probability $\prod_{k \neq j} \mu_k(s_k)$.

Finally, choose some arbitrary player $k \notin \{i, j\}$. Since t_i believes that k has CIB, and believes that k's type is t_k, type t_k must have CIB. Hence,

$$\beta_k(t_k)(s_{-k}|t_{-k}) = \prod_{l \neq k} \beta_k(t_k)(s_l|t_l) \tag{4.3}$$

for every $s_{-k} \in S_{-k}$. Let $l \notin \{i, k\}$. By Lemma 4.2 we know that t_i has PB. Since t_i has PB, and believes that k's type is t_k, type t_i's belief about l's strategy choice must be equal to k's belief about l's strategy choice, and hence $\beta_k(t_k)(s_l|t_l) = \beta_i(t_i)(s_l|t_l) = \mu_l(s_l)$ for every $s_l \in S_l$. Since t_i believes that j has PB, and believes that j has type t_j, type t_j must have PB. Hence, t_j's belief about i's strategy choice must be equal to k's belief about i's strategy choice, which implies that $\beta_k(t_k)(s_i|t_i) = \beta_j(t_j)(s_i|t_i) = \mu_i(s_i)$. Combined with (4.3) we obtain that t_k believes that every opponents' strategy profile s_{-k} is chosen with probability $\prod_{l \neq k} \mu_l(s_l)$. So, t_i believes that every $k \neq j$ believes that every opponents' strategy profile s_{-k} is chosen with probability $\prod_{l \neq k} \mu_l(s_l)$. Hence, we have shown property (2).

Since we already know from Lemma 4.1 that t_i believes that every opponent j believes that i's type is t_i, property (3) follows immediately. This completes the proof. □

We are now ready to prove our main theorem.

Theorem 4.5 *Let $\mathcal{M} = (T_i, \beta_i)_{i \in I}$ be a finite epistemic model for a finite game Γ. Let $t_i \in T_i$ be a type that satisfies BOR, SRB and CIB, and believes that every opponent satisfies BOR, SRB, PB and CIB. Then, every strategy that is rational for t_i is a Nash strategy for Γ.*

Proof Let t_i be a type that satisfies BOR, SRB and CIB, and believes that every opponent satisfies BOR, SRB, PB and CIB. Let $\mu_k \in \Delta(S_k)$, $k \in I$, be the probability distributions obtained from Lemma 4.4. We show that $(\mu_k)_{k \in I}$ is a Nash equilibrium.

Suppose first that $j \neq i$ and that $\mu_j(s_j) > 0$ for some $s_j \in S_j$. Since t_i believes in j's rationality, and, by (2), believes that j's belief about the opponents' strategy choices is given by $\mu_{-j} = (\mu_k)_{k \neq j}$, it follows that s_j must be optimal against μ_{-j}. Finally, let $\mu_i(s_i) > 0$ for some $s_i \in S_i$. Choose some arbitrary opponent j. Since, by (2), t_i's belief about j's belief about the opponents' strategy choice is given by μ_{-j}, type t_i believes that j believes that i chooses s_i with positive probability. Since t_i believes that j believes in i's rationality, and since, by (3), t_i's belief about j's belief about i's belief about the opponents' strategy choices is given by μ_{-i}, it follows that s_i must be optimal against μ_{-i}. Hence, $(\mu_j)_{j \in I}$ is a Nash equilibrium.

Now, choose some strategy s_i that is rational for t_i. Since, by (1), t_i's belief about the opponents' strategy choices is μ_{-i}, strategy s_i must be optimal against μ_{-i}. As (μ_i, μ_{-i}) is a Nash equilibrium, s_i is a Nash strategy. This completes the proof. □

Theorem 4.5 thus provides sufficient doxastic conditions for choosing Nash *strategies*. The question remains whether the conditions in this theorem also lead, in some sense, to Nash *equilibrium* in beliefs. The answer is "yes." Consider, namely, a type t_i that satisfies BOR, SRB and CIB, and believes that every opponent satisfies BOR, SRB, PB and CIB. In the proof of Theorem 4.5, we have shown that there exist probability distributions $\mu_k \in \Delta(S_k)$, $k \in I$, such that (1) for $j \neq i$, μ_j is t_i's belief about j's strategy choice, and also t_i's belief about any other player's belief about player j's strategy choice, and (2) μ_i is t_i's belief about any other player's belief about his own strategy choice. Moreover, we have shown in the proof that the profile $(\mu_k)_{k\in I}$ is a Nash equilibrium in beliefs. We thus obtain the following corollary.

Corollary 4.6 *Let $\mathcal{M} = (T_i, \beta_i)_{i\in I}$ be a finite epistemic model for a finite game Γ. Let $t_i \in T_i$ be a type that satisfies BOR, SRB and CIB, and believes that every opponent satisfies BOR, SRB, PB and CIB. For every player $j \neq i$, let $\mu_j \in \Delta(S_j)$ be t_i's belief about j's strategy choice, and let μ_i be t_i's belief about any other player's belief about his own strategy choice. Then, $(\mu_j)_{j\in I}$ is a Nash equilibrium in beliefs.*

Our last result shows that every Nash strategy can be chosen rationally by a type that satisfies BOR, SRB and CIB, and believes that his opponents satisfy BOR, SRB, PB and CIB. In particular, the combination of these seven events is shown to be possible.

Theorem 4.7 *Let Γ be a finite game, and let s_i be a Nash strategy for player i in Γ. Then, there exists a finite epistemic model $\mathcal{M} = (T_i, \beta_i)_{i\in I}$ for Γ, and a type $t_i \in T_i$, such that s_i is rational for t_i, type t_i satisfies BOR, SRB, and CIB, and believes that every opponent satisfies BOR, SRB, PB and CIB.*

Proof Let s_i be a Nash strategy for Γ. Then, there is a Nash equilibrium $(\mu_j)_{j\in I}$ for Γ such that s_i is optimal for i against μ_{-i}. We define the epistemic model $\mathcal{M} = (T_i, \beta_i)_{i\in I}$ as follows: Define $T_j := \{\hat{t}_j\}$ for every player j, and let $\beta_j(\hat{t}_j)$ be the probability distribution on $S_{-j} \times T_{-j}$ given by

$$\beta_j(\hat{t}_j)(s_{-j}, t_{-j}) := \begin{cases} \prod_{k\neq j} \mu_k(s_k), & \text{if } t_{-j} = \hat{t}_{-j} \\ 0, & \text{otherwise.} \end{cases}$$

We show that for every player j, the type $\hat{t}_j$ satisfies BOR, SRB, PB and CIB. Since $\hat{t}_i$ believes, for every opponent j, that j's type is $\hat{t}_j$, it would follow that $\hat{t}_i$ believes that every opponent satisfies these four conditions as well.

BOR: Let k be an opponent for j, and let s_k be a strategy for k with $\beta_j(\hat{t}_j)(s_k, \hat{t}_k) > 0$. Then, $\mu_k(s_k) > 0$. Since $(\mu_j)_{j\in I}$ is a Nash equilibrium, s_k is optimal for k against μ_{-k}, and hence s_k is rational for $\hat{t}_k$. Therefore, $\hat{t}_j$ satisfies BOR.

SRB: By construction, $\hat{t}_j$ believes that every opponent k has type $\hat{t}_k$. Moreover, every such type $\hat{t}_k$ believes that j's type is $\hat{t}_j$. Hence, $\hat{t}_j$ believes that every opponent believes that j's type is $\hat{t}_j$. Now, suppose that $\hat{t}_j$ believes an event E with probability p. Then, since $\hat{t}_j$ believes that every opponent believes that j's type is $\hat{t}_j$, type $\hat{t}_j$ believes that every opponent believes that j believes E with probability p. Consequently, $\hat{t}_j$ satisfies SRB.

PB: Let k, l be two different opponents for j, let $E_k \subseteq S_k \times T_k$, and suppose that $\hat{t}_j$ believes E_k with probability p. By construction, $\hat{t}_j$ believes that l's type is $\hat{t}_l$, and $\hat{t}_l$'s belief about player k is the same as $\hat{t}_j$'s belief about player k. Hence, $\hat{t}_j$ believes that l believes E_k with probability p. We may thus conclude that $\hat{t}_j$ satisfies PB.

CIB: It follows immediately from the construction of the epistemic model that $\hat{t}_j$ satisfies CIB.

Hence, we may conclude that every $\hat{t}_j$ satisfies BOR, SRB, PB and CIB. This implies that $\hat{t}_i$ satisfies these four conditions, and believes that every opponent satisfies these four conditions too. Recall that the Nash equilibrium $(\mu_j)_{j \in I}$ was chosen such that s_i is optimal for i against μ_{-i}. Since μ_{-i} is $\hat{t}_i$'s belief about the opponents' strategy choices, s_i is rational for $\hat{t}_i$. This completes the proof. □

5 No conditions can be dropped

In Theorem 4.5 we have shown that the conditions BOR, SRB, CIB, and belief in BOR, SRB, PB and CIB, lead to Nash strategy choices. So, in total we have seven conditions that we impose on a player's beliefs. In this section we prove that this result no longer holds if we drop one of these seven conditions. In particular, we show that these seven conditions are logically independent.

Dropping BOR: Consider the two-player game in Fig. 4, where player 1 chooses the rows and player 2 the columns. Construct an epistemic model $\mathcal{M}$ such that $T_1 = \{t_1\}$, $T_2 = \{t_2\}$, $\beta_1(t_1)$ assigns probability 1 to (d, t_2), and $\beta_2(t_2)$ assigns probability 1 to (a, t_1). Then, t_1 does not BOR, since d is not optimal for t_2. On the other hand, t_1 believes that player 2 BOR, since a is optimal for t_1. Type t_1 has SRB since t_1 believes that player 2 believes that his type is t_1. Similarly, t_2 has SRB, and hence t_1 believes in SRB. Clearly, t_1 has CIB, and believes that player 2 has PB and CIB, since these conditions are automatically satisfied for two players. Hence, t_1 does not BOR, but satisfies the other six conditions. Strategy a is rational for t_1, but a is not a Nash strategy.

Dropping belief in BOR: Consider again the two-player game in Fig. 4. Construct an epistemic model $\mathcal{M}$ such that $T_1 = \{t_1\}$, $T_2 = \{t_2\}$, $\beta_1(t_1)$ assigns probability 1 to (d, t_2), and $\beta_2(t_2)$ assigns probability 1 to (c, t_1). Then, t_1 does not believe in BOR, since c is not optimal for t_1. However, t_1 satisfies the other six conditions. Strategy a is optimal for t_1, but a is not a Nash strategy.

	d	e
a	2, 0	1, 1
b	1, 0	2, 1
c	0, 1	0, 0

Fig. 4 BOR and belief in BOR cannot be dropped

	d	e	f
a	3, 0	0, 3	0, 0
b	0, 3	3, 0	0, 0
c	2, 0	2, 0	2, 2

Fig. 5 SRB and belief in SRB cannot be dropped

g	d	e	f
a	3, 3, 0	3, 0, 3	0, 2, 0
b	0, 0, 0	0, 0, 0	0, 2, 0
c	2, 0, 0	2, 0, 0	2, 2, 0

h	d	e	f
a	0, 0, 0	0, 3, 3	0, 2, 0
b	0, 0, 0	3, 3, 0	0, 2, 0
c	2, 0, 0	2, 0, 0	2, 2, 0

i	d	e	f
a	0, 0, 2	0, 0, 2	0, 2, 2
b	0, 0, 2	0, 0, 2	0, 2, 2
c	2, 0, 2	2, 0, 2	2, 2, 2

Fig. 6 Belief in PB cannot be dropped

Dropping SRB: Consider the two-player game in Fig. 5. It can be shown that (c, f) is the only Nash equilibrium, and hence c is the only Nash strategy for player 1. Construct an epistemic model $\mathcal{M}$ such that $T_1 = \{t_1^a, t_1^b\}$, $T_2 = \{t_2\}$, $\beta_1(t_1^a)$ assigns probability 1 to (d, t_2), $\beta_1(t_1^b)$ assigns probability 1 to (e, t_2), and $\beta_2(t_2)$ assigns probability 1/2 to (a, t_1^a) and probability 1/2 to (b, t_1^b). Then, t_1^a does not have SRB, since t_1^a believes that player 2 believes with probability 1/2 that his type is t_1^b and not t_1^a. However, t_1^a satisfies the other six conditions. Strategy a is optimal for t_1^a, but a is not a Nash strategy.

Dropping belief in SRB: Consider again the two-player game in Fig. 5. Recall that (c, f) is the only Nash equilibrium, and hence f is the only Nash strategy for player 2. Construct an epistemic model $\mathcal{M}$ such that $T_1 = \{t_1^a, t_1^b\}$, $T_2 = \{t_2\}$, $\beta_1(t_1^a)$ assigns probability 1 to (d, t_2), $\beta_1(t_1^b)$ assigns probability 1 to (e, t_2), and $\beta_2(t_2)$ assigns probability 1/2 to (a, t_1^a) and probability 1/2 to (b, t_1^b). Then, t_2 does not believe in SRB, since t_1^a and t_1^b do not have SRB. However, t_2 satisfies the other six conditions. Strategy d is optimal for t_2, but d is not a Nash strategy.

Dropping belief in PB: Consider the three-player game in Fig. 6. Here, player 1 chooses the row, player 2 chooses the column, and player 3 chooses the matrix (g, h or i). Construct an epistemic model $\mathcal{M}$ such that $T_1 = \{t_1\}$, $T_2 = \{t_2\}$, $T_3 = \{t_3\}$, $\beta_1(t_1)$ assigns probability 1 to $((e, t_2), (g, t_3))$, $\beta_2(t_2)$ assigns probability 1 to $((a, t_1), (h, t_3))$, and $\beta_3(t_3)$ assigns probability 1 to $((a, t_1), (e, t_2))$. Type t_3 does not believe in PB, since t_1 does not have PB. In order to see this, note that t_1 believes that player 3 chooses g, whereas t_1 believes that player 2 believes that player 3 chooses h. It can easily be verified that t_3 satisfies the other six conditions.

Strategy g is rational for type t_3. However, we will show that g is not a Nash strategy. Suppose, contrary to what we want to show, that g would be a Nash strategy. Then, there would be a Nash equilibrium (μ_1, μ_2, μ_3) such that g would be optimal against (μ_1, μ_2). Strategy g can only be optimal against (μ_1, μ_2) if $\mu_1(a) > 0$ and $\mu_2(e) > 0$. Since (μ_1, μ_2, μ_3) is a Nash equilibrium, this implies that a is optimal against (μ_2, μ_3) and e is optimal against (μ_1, μ_3). This, in turn, implies that $\mu_3(g) \geq 2/3$ and $\mu_3(h) \geq 2/3$, which is clearly impossible. Hence, g is not a Nash strategy.

g	d	e	f
a	$0,0,3$	$0,0,3$	$3,3,3$
b	$0,0,0$	$0,0,0$	$3,3,3$
c	$3,3,3$	$3,3,3$	$3,3,3$

h	d	e	f
a	$0,0,2$	$0,0,0$	$0,0,0$
b	$0,0,0$	$0,0,2$	$0,0,0$
c	$0,0,0$	$0,0,0$	$0,0,0$

i	d	e	f
a	$0,0,0$	$0,0,0$	$3,3,3$
b	$0,0,3$	$0,0,3$	$3,3,3$
c	$3,3,3$	$3,3,3$	$3,3,3$

Fig. 7 CIB and belief in CIB cannot be dropped

Dropping CIB: Consider the three-player game in Fig. 7. Construct an epistemic model $\mathcal{M}$ such that $T_1 = \{t_1\}$, $T_2 = \{t_2\}$, $T_3 = \{t_3\}$, $\beta_1(t_1)$ assigns probability $1/2$ to $((d, t_2), (h, t_3))$ and probability $1/2$ to $((e, t_2), (h, t_3))$, $\beta_2(t_2)$ assigns probability $1/2$ to $((a, t_1), (h, t_3))$ and probability $1/2$ to $((b, t_1), (h, t_3))$, and $\beta_3(t_3)$ assigns probability $1/2$ to $((a, t_1), (d, t_2))$ and probability $1/2$ to $((b, t_1), (e, t_2))$. Then, t_3 does not have CIB. It may be verified that t_3 satisfies the other six conditions.

Strategy h is rational for t_3. However, we will show that h is not a Nash strategy. Assume, namely, that (μ_1, μ_2, μ_3) would be a Nash equilibrium such that h would be optimal against (μ_1, μ_2). If $\mu_1(a) \leq 1/2$, it can be shown that $u_3(h, \mu_1, \mu_2) < u_3(i, \mu_1, \mu_2)$. If $\mu_1(a) \geq 1/2$, it can be shown that $u_3(h, \mu_1, \mu_2) < u_3(g, \mu_1, \mu_2)$. Hence, h can never be optimal against (μ_1, μ_2), and therefore h is not a Nash strategy.

Dropping belief in CIB: Consider again the three-player game in Fig. 7. Construct an epistemic model $\mathcal{M}$ such that $T_1 = \{t_1\}$, $T_2 = \{t_2\}$, $T_3 = \{t_3\}$, $\beta_1(t_1)$ assigns probability $1/2$ to $((d, t_2), (h, t_3))$ and probability $1/2$ to $((e, t_2), (h, t_3))$, $\beta_2(t_2)$ assigns probability $1/2$ to $((a, t_1), (h, t_3))$ and probability $1/2$ to $((b, t_1), (h, t_3))$, and $\beta_3(t_3)$ assigns probability $1/2$ to $((a, t_1), (d, t_2))$ and probability $1/2$ to $((b, t_1), (e, t_2))$. Then, t_1 does not believe in CIB, since t_3 does not have CIB. It may be verified that t_1 satisfies the other six conditions.

Strategy a is rational for t_1. However, we will show that a is not a Nash strategy. We have seen above that h is not a Nash strategy. In particular, there is no Nash equilibrium that assigns positive probability to h. But then, a cannot be optimal in a Nash equilibrium, and hence a is not a Nash strategy.

6 Discussion of main result

Our main result comes in two parts: In Theorem 4.5 we first show that, if a player satisfies BOR, SRB and CIB, and believes that every opponent satisfies BOR, SRB, PB and CIB, then this player will eventually choose a Nash strategy. We next prove in Theorem 4.7 that every Nash strategy for player i can be supported by a player i belief hierarchy that satisfies these seven conditions. By combining these two results, we may thus view the seven conditions above as a one-person doxastic characterization of Nash strategies.

6.1 Discussion of SRB and PB

Since SRB and PB are new, and play a crucial role in the doxastic characterization of Nash strategies, we will discuss these two conditions in some more depth here. Both SRB and PB reflect the event that player i uses his own beliefs as a “focal point” for what he believes that others believe. In SRB, namely, player i uses his own belief about himself (assuming that he has a correct belief about himself) as a focal point for what he believes that others believe about himself, whereas in PB player i uses his own belief about opponent j as a focal point for what he believes that others believe about opponent j. Now, can it be reasonable, at an intuitive level, to take your own belief as a focal point for what you believe that others believe? In my opinion, this can be reasonable if there are no serious hints that some of your opponents really reason differently than you do. This may happen if you have little or no information about the opponents' characteristics, and in particular about their ways of reasoning. In such situations, one "easy" way to form a belief about what others believe is simply to assume that your opponents have the same beliefs as you do. Moreover, if your own belief satisfies certain properties, such as BOR, then by believing that others have the same belief as you do, you will be sure to believe that others will share these properties too. Hence, if you face a serious lack of information about your opponents' beliefs, then choosing your own belief as a "focal candidate" for the beliefs of others may make intuitive sense. This lack of information about others will often occur in the type of setting we consider, namely static games in which the game is only played once, and players do not communicate with each other before reaching a decision.

6.2 Can SRB be weakened?

In Sect. 5 we have seen that none of the seven conditions in Theorem 4.5 can be dropped completely. In particular, by completely dropping the condition that player i has SRB, or believes that every opponent has SRB, one can no longer guarantee that player i will choose a Nash strategy. However, can we replace the notion of SRB in Theorem 4.5 by a weaker requirement without violating the content of the theorem? The answer is "yes." Consider, namely, the weaker condition stating that t_i believes that every opponent holds a correct belief about player i's belief about his opponents' *choices*. Let us call this condition *weakly self-referential beliefs* (WSRB). The difference with SRB is that the latter requires that t_i believes that every opponent holds a correct belief about player i's *complete belief hierarchy*, not only about his belief about the opponents' *choices*. One can show the following result by basically copying the proof from Theorem 4.5.

Theorem 6.1 *Let $\mathcal{M} = (T_i, \beta_i)_{i \in I}$ be a finite epistemic model for a finite game Γ. Let $t_i \in T_i$ be a type that satisfies BOR, WSRB and CIB, and believes that every opponent satisfies BOR, WSRB, PB and CIB. Then, every strategy that is rational for t_i is a Nash strategy for Γ.*

6.3 Single player's perspective

Remember that our Theorem 4.5 only imposes conditions on the beliefs of a single player, namely player i. These conditions guarantee that player i will choose a Nash strategy. Moreover, as we have seen in Corollary 4.6, the same conditions imply that player i's belief about his opponents' choices, together with his belief about the opponents' belief about his own choice, constitute a Nash equilibrium. In this context, a

	c	d
a	1, 1	0, 0
b	0, 0	1, 1

Fig. 8 Beliefs of different players may not constitute a Nash equilibrium

Nash equilibrium is interpreted as a "personal object" that is completely inside player i's mind.

Now, suppose there are only two players in a game, player i and player j, who both satisfy the conditions in Theorem 4.5. Do the players' beliefs about the opponent's choice then constitute a Nash equilibrium? The answer is "no." Although the conditions guarantee that player i's personal beliefs about player j constitute a Nash equilibrium (in the sense of Corollary 4.6) and that player j's personal beliefs about player i constitute a Nash equilibrium, it may well be that both players have *different* Nash equilibria in their minds. Consequently, i's belief about j's choice and j's belief about i's choice may not constitute a Nash equilibrium in beliefs. Consider, for instance, the coordination game in Fig. 8. Construct an epistemic model $\mathcal{M}$ such that $T_1 = \{t_1^a, t_1^b\}$, $T_2 = \{t_2^c, t_2^d\}$, $\beta_1(t_1^a)$ assigns probability 1 to (c, t_2^c), $\beta_1(t_1^b)$ assigns probability 1 to (d, t_2^d), $\beta_2(t_2^c)$ assigns probability 1 to (a, t_1^a) and $\beta_2(t_2^d)$ assigns probability 1 to (b, t_1^b). Then, every type in T_1 and T_2 satisfies the conditions in Theorem 4.5, however t_1^a's belief about 2's choice and t_2^d's belief about 1's choice do not constitute a Nash equilibrium.

In this respect, our Theorem 4.5 differs crucially from Aumann and Brandenburger (1995), Asheim (2006) and Brandenburger and Dekel (1987) who impose conditions on the beliefs of *all* players which, together, imply that the players' beliefs about the opponents' choices constitute a Nash equilibrium. In the following section, we will explore the differences with these models in some more detail.

7 Comparison with other models

7.1 Aumann and Brandenburger's model

Aumann and Brandenburger (1995) (AB from now on) make a distinction between the case of two players and the case of more than two players, and provide sufficient conditions for Nash equilibrium for both cases. In AB's model, a type for player i does not only specify i's belief hierarchy, but also i's strategy choice and i's utility function. It is therefore possible to say that type t_i is rational. AB's theorem for two-player games may be formulated as follows: Consider a pair (u_1, u_2) of utility functions, a pair $(\mu_1, \mu_2) \in \Delta(S_1) \times \Delta(S_2)$ of probability distributions over strategy choices, and a pair (t_1, t_2) of types. If at (t_1, t_2) it is true that (1) both players are rational and believe that the opponent is rational, (2) both players i have utility function u_i and believe that opponent j has utility function u_j, and (3) both players i have belief μ_j

about j's strategy choice, and believe that j has belief μ_i about i's strategy choice, then (μ_1, μ_2) is a Nash equilibrium with respect to (u_1, u_2).

An important difference with our model is that AB simultaneously impose conditions on the belief hierarchies of *both* players, whereas our model only imposes conditions on the beliefs of a *single* player. So, in this sense AB's model is more restrictive. However, if we concentrate on the conditions that AB impose on the beliefs of a *single* player, we see that these are weaker than our conditions in Theorem 4.5 for two-player games. For instance, condition (1) above implies that player i BOR, but not necessarily that i believes that j BOR, as we impose in Theorem 4.5. Also, condition (3) above does not imply that i has SRB: i may have belief μ_j about j's choice, and at the same time believe that j believes that i has a different belief about j's choice. So, in fact, condition (3) does not even imply that i has WSRB (see Theorem 6.1). Hence, when analyzed from a single player's perspective only, AB's conditions are weaker than our conditions for two-player games.

In order to highlight the differences with our model, consider again the game from Fig. 8. In terms of AB's model, let $T_1 = \{t_1^a, \hat{t}_1^a\}$ and $T_2 = \{t_2^c, \hat{t}_2^c, t_2^d\}$ be sets of types where t_1^a, $\hat{t}_1^a$ choose a, t_2^c and $\hat{t}_2^c$ choose c, t_2^d chooses d, t_1^a believes that 2's type is $\hat{t}_2^c$, $\hat{t}_1^a$ believes that 2's type is t_2^d, t_2^c believes that 1's type is t_1^a, $\hat{t}_2^c$ believes that 1's type is $\hat{t}_1^a$ and t_2^d believes that 1's type is t_1^a. See Fig. 9 for an illustration. Here, the arrows denote beliefs. For instance, the arrow from t_1^a to $\hat{t}_2^c$ means that type t_1^a believes that player 2's type is $\hat{t}_2^c$. Furthermore, assume that each type has the utility function as depicted in Fig. 8. Let μ_1 be the probability distribution that assigns probability 1 to a, and let μ_2 be the probability distribution that assigns probability 1 to c. Then at (t_1^a, t_2^c) conditions (1), (2) and (3) above are satisfied. That is, at (t_1^a, t_2^c) AB's sufficient conditions for Nash equilibrium are met. However, type t_1^a does not believe that player 2 BOR. Namely, t_1^a believes that player 2 believes that player 1 is of type $\hat{t}_1^a$ and chooses a. Type $\hat{t}_1^a$, in turn, believes that player 2 chooses d, and hence a is not rational for type $\hat{t}_1^a$. So, t_1^a believes that player 2 believes that player 1 chooses irrationally. Also, t_1^a does not have SRB: t_1^a believes that player 2 chooses c, whereas t_1^a believes that player 2 believes that player 1 is of type $\hat{t}_1^a$ who believes that player 2 chooses d. In fact, t_1^a does not even have WSRB. Hence, type t_1^a does not satisfy our sufficient conditions for Nash strategy choice, although AB's conditions are met at (t_1^a, t_2^c).

The example above shows, in particular, that AB's conditions do not imply common belief in rationality at (t_1, t_2). Here, by common belief in rationality, at (t_1, t_2) we mean that both players are rational at (t_1, t_2), both players believe at (t_1, t_2) that both players are rational, and so on. Polak (1995) proves that, if conditions (2) and (3)

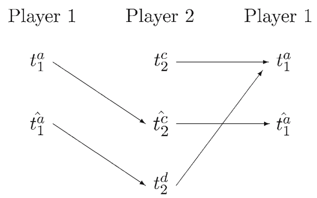

Fig. 9 An example illustrating AB's theorem

above are strengthened by imposing *common* belief in the conjectures (μ_1, μ_2) and the utility functions (u_1, u_2), then common belief in rationality will hold at (t_1, t_2). Brandenburger and Dekel (1989) have shown a related result for two-player games: If there is common belief in the conjectures (μ_1, μ_2) and the utility functions (u_1, u_2), then (μ_1, μ_2) is a Nash equilibrium if and only if common belief in rationality holds.

For the case of more than two players, AB add the following two conditions: (4) at the type profile $(t_i)_{i\in I}$ the types' belief hierarchies are derived from a common prior probability distribution on the set of type profiles, and (5) at $(t_i)_{i\in I}$ there is common belief in the profile $(\beta_i)_{i\in I} \in \times_{i\in I}\Delta(S_{-i})$ of beliefs about the opponents' strategies. By the latter we mean that t_i's belief about the opponents' choices is β_i, that t_i believes that every opponent j has belief β_j about the other players' choices, that t_i believes that every opponent j believes that every other player k has belief β_k about the opponents' choices, and so on. By adding these conditions (4) and (5), AB are able to show that there is some profile $(\mu_i)_{i\in I} \in \times_{i\in I}\Delta(S_i)$ of probability distributions over strategy choices such that for every player i, type t_i's belief about the opponents' choices is given by $(\mu_j)_{j\neq i}$, and that $(\mu_i)_{i\in I}$ is a Nash equilibrium with respect to $(u_i)_{i\in I}$. In particular, the conditions (4) and (5) imply that i's belief about j's choice is stochastically independent from i's belief about k's choice, and that two different players i and j have the same belief about k's choice. These two properties are crucial for their proof. In our model, these two properties follow from the assumption that t_i has SRB and CIB, and believes that every opponent has SRB, PB and CIB (see Lemma 4.4). Hence, one could say that in our model the conditions of CIB and belief in PB and CIB play a similar role as the conditions (4) and (5) in AB.

7.2 Asheim's model

Asheim (2006, p. 5), provides a sufficient condition for Nash equilibrium for the case of two players. He basically uses the same epistemic model as we do, and his result may be stated as follows: Consider a pair (t_1, t_2) of types, and assume that (1) t_1 and t_2 BOR, and (2) t_1 believes that 2's type is t_2, and t_2 believes that 1's type is t_1. Then, t_1's belief about 2's choice and t_2's belief about 1's choice constitute a Nash equilibrium.

Similar to AB, Asheim simultaneously imposes conditions on the belief hierarchies of *both* players. However, if we analyze Asheim's conditions from a single-player perspective, then we see that Asheim's conditions imply ours, Focus, namely, on a single player, say player i. Conditions (1) and (2) above together imply that t_i BOR, and that t_i believes that j BOR, since t_i believes that j's type is t_j, who is supposed to BOR. Moreover, conditions (1) and (2) imply that t_i has SRB and believes that j has SRB. Namely, t_i believes that j's type is t_j, who believes that i's type is t_i. Hence, t_i believes that j believes that i's type is t_i, and hence t_i has SRB. Also, by the same reasoning, t_j has SRB. Since t_i believes that j's type is t_j, type t_i believes that j has SRB. Summarizing, we may conclude that Asheim's conditions (1) and (2) imply that t_i BOR, believes that j BOR, has SRB, and believes that j has SRB. So, Asheim's conditions, when analyzed from a single-player's perspective, imply our sufficient conditions for Nash strategy choice.

Tan and Werlang (1988), in their Theorem 6.2.1, p. 382, provide conditions for Nash equilibrium in two-player games that are similar to Asheim's, although somewhat stronger. They assume, like Asheim, that t_1 believes that 2's type is t_2, and that t_2 believes that 1's type is t_1. However, Tan and Werlang impose *common* belief in rationality at (t_1, t_2), instead of only requiring that t_1 and t_2 BOR.

7.3 Brandenburger and Dekel's model

Brandenburger and Dekel (1987) (BD from now on) use a model which substantially differs from ours and the ones above. BD assume that there is a finite state space Ω, that every player i holds a prior belief $P_i \in \Delta(\Omega)$, that for every player i there is an information partition $\mathcal{H}_i$ of Ω, and that for every $H_i \in \mathcal{H}_i$ there is a conditional belief $P_i(\cdot|H_i) \in \Delta(H_i)$ which is derived from P_i by Bayes' rule whenever possible. A strategy map for player i is an $\mathcal{H}_i$-measurable map f_i from Ω to S_i. A profile $(f_i)_{i\in I}$ of strategy maps is called an *a posteriori equilibrium* if for every player i, at every $H_i \in \mathcal{H}_i$ the prescribed strategy is optimal given the conditional belief $P_i(\cdot|H_i)$ and the opponents' strategy maps $(f_j)_{j\neq i}$. Hence, it is implicitly assumed that players have correct beliefs about the opponents' strategy maps, and that there is common belief in this event. The prior beliefs $(P_i)_{i\in I}$ are called *concordant* if, for every choice of the strategy maps, two different players i and j have the same prior belief about k's choice. The information partitions $(\mathcal{H}_j)_{j\neq i}$ are called P_i*-conditionally independent* if, for every choice of the strategy maps, i's belief about j'choice is independent from i's belief about k's choice. The information partitions $(\mathcal{H}_j)_{j\neq i}$ are called P_i*-informationally independent* if, for every choice of the strategy maps, i's belief about the opponents' choices does not depend on the information set H_i.

BD's theorem on page 1401 can now be stated as follows: suppose that the prior beliefs $(P_i)_{i\in I}$ and the information partitions $(\mathcal{H}_i)_{i\in I}$ are such that (1) the prior beliefs $(P_i)_{i\in I}$ are concordant, and for every i, the information partitions $(\mathcal{H}_j)_{j\neq i}$ are (2) P_i-conditionally independent and (3) P_i-informationally independent. Then, for every a posteriori equilibrium there is a profile $(\mu_i)_{i\in I} \in \times_{i\in I}\Delta(S_i)$ of probability distributions over strategy choices such that, for every i, player i's prior belief about the opponents' choices is given by $(\mu_j)_{j\neq i}$, and $(\mu_i)_{i\in I}$ is a Nash equilibrium.

In BD's theorem, selecting an a posteriori equilibrium implies that player i BOR. Concordance of the prior beliefs guarantees that i and j have the same beliefs about k's choice, and therefore implies our condition of PB. The assumption in BD that the information partitions $(\mathcal{H}_j)_{j\neq i}$ are P_i-informationally independent guarantees that i's belief about his opponents is independent of his information state, and therefore i believes that his opponents are right about his own beliefs. (Recall the implicit assumption that players have correct beliefs about the opponents' strategy maps, and that there is common belief in this event). The assumption of P_i-informational independence therefore implies that player i has SRB. In BD, P_i-conditional independence guarantees that i's belief about j's choice is independent from i's belief about k's choice. Since P_j- and P_k-informational independence implies that i only deems one belief hierarchy, and hence only one type, possible for opponents j and k, P_i-conditional independence and P_j- and P_k-informational independence together imply that player i has CIB. Summarizing, we see that BD's conditions imply that player i BOR, and has SRB, PB and CIB. Now, fix an opponent j. By BD's conditions, player j also BOR, and has SRB, PB and CIB. Since BD's condition of P_j-informational independence implies that player i is correct about j's belief hierarchy, it follows that i believes that every opponent j BOR, and has SRB, PB and CIB. We may thus conclude that BD's conditions, when analyzed from a single-player's perspective, imply ours.

Acknowledgements I wish to thank Geir Asheim and two referees for their helpful suggestions.

References

Asheim, G. B. (2006). *The consistent preferences approach to deductive reasoning in games*. Theory and Decision Library. Dordrecht: Springer.

Aumann, R., & Brandenburger, A. (1995). Epistemic conditions for Nash equilibrium. *Econometrica, 63*, 1161–1180.

Bernheim, D. (1984). Rationalizable strategic behavior. *Econometrica, 52*, 1007–1028.

Brandenburger, A., & Dekel, E. (1987). Rationalizability and correlated equilibria. *Econometrica, 55*, 1391–1402.

Brandenburger, A., & Dekel, E. (1989). The role of common knowledge assumptions in game theory. In F. Hahn (Ed.), *The economics of missing markets, information and games* (pp. 46–61). Oxford: Oxford University Press.

Brandenburger, A., & Friedenberg, A. (2006). Intrinsic correlation in games. Available at http://pages.stern.nyu.edu/ abranden/

Myerson, R. (1991). *Game theory: Analysis of conflict*. Cambridge, Massachusetts: Harvard University Press.

Nash, J. (1951). Non-cooperative games. *Annals of Mathematics, 54*, 286–295.

Pearce, D. (1984). Rationalizable strategic behavior and the problem of perfection. *Econometrica, 52*, 1029–1050.

Polak, B. (1995). Epistemic conditions for Nash equilibrium, and common knowledge of rationality. *Econometrica, 67*, 673–676.

Tan, T., & Werlang, S. R. C. (1988). The bayesian foundations of solution concepts of games. *Journal of Economic Theory, 45*, 370–391.