

# WEAKLY DOMINATED STRATEGIES: A MYSTERY CRACKED

DOV SAMET

ABSTRACT. An informal argument shows that common knowledge of rationality implies the iterative elimination of strongly dominated strategies. Rationality here means that players do not play strategies that are strongly dominated relative to their knowledge. We formalize and prove this claim. When by rationality we mean that players do not play strategies that are weakly dominated relative to their knowledge, then common knowledge of rationality does not imply iterative elimination of weakly dominated strategies. We show that it does imply an iterative elimination of weak-dominance *bases*. The iterative elimination of strong-dominance bases coincides with the iterative elimination of strongly dominated strategies.

## 1. INTRODUCTION

Iterative elimination of strongly dominated strategies is usually justified by the assumption that it is common knowledge that players are *strong-dominance rational*, by which we mean that they do not play strategies that are strongly dominated given their knowledge about the opponents' strategies.<sup>1</sup> The argument for this in the literature is informal, as knowledge, and a fortiori strong-dominance rationality which is defined in terms of knowledge, are not fully formalized. We show that the argument can be made rigorous by formally modeling knowledge and strong-dominance rationality.

The picture changes when weakly dominated strategies are considered. Iterative elimination of weakly dominated strategies fails to capture common knowledge of *weak-dominance rationality*, which requires that players do not play strategies that are weakly dominated given their knowledge about the opponents' strategies. The problem does not lie in the assumption of weak-dominance rationality, but in the process of elimination. In a nutshell, the problem is that strategies that are eliminated in early stages of the process, given the knowledge of the players at these stages, may not be weakly dominated given the knowledge at the end of the process, which is the knowledge players have, according to this argument, under the assumption of common knowledge of weak-dominance rationality.

Despite the awareness of this problem, no suggestion has been made how to fix the process of iterated elimination of weakly dominated strategies in order to capture common knowledge of weak-dominance rationality, due to the lack of formalization of weak-dominance rationality. Here, we formalize this notion analogously to the notion of strong-dominance rationality, and show that it implies an iterative process of elimination of *dominance bases* of weakly dominated strategies. This process circumvents the problem of iterative elimination of weakly dominated

---

*Date:* Draft of August 29, 2013.

<sup>1</sup>See, for example, Myerson (1991, 57–61) and Binmore (1991, 149–150).

strategies alluded above. A dominance base of a dominated strategy of a player is a combination of this strategy with strategies of her opponents, which has a strong inequality in the set of inequalities that describe the dominance relationship.

Iterative elimination of dominance bases is the process that expresses common knowledge of both strong-dominance and weak-dominance rationality. When dominance bases of strongly dominated strategies are eliminated iteratively, the result is the same as the iterative elimination of strongly dominated strategies. In the case of weak dominance, iterative elimination of dominance bases is different from the iterative elimination of weakly dominated strategy, and unlike the latter, it reflects common knowledge of weak-dominance rationality.

Hillas and Samet (2013) study strong-dominance and weak-dominance rationality under the names weak rationality and strong rationality correspondingly, and show their relation to correlated equilibrium. Here we use these notions to clarify the nature of the age-old solutions of iterative elimination of dominated strategies. We propose the iterative elimination of dominance bases which coincides with the known solution for the strong dominance case, and provides an alternative, contradiction free, replacement of the solution for the weak dominance case.

## 2. ITERATIVE ELIMINATION OF DOMINATED STRATEGIES

Let  $G$  be a game with a finite set of players  $I$ , and a finite set of strategies  $S_i$  for each player  $i$ . The set of strategy profiles is  $S = \times_i S_i$ , and the set of the profiles of  $i$ 's opponents is  $S_{-i} = \times_{j \neq i} S_j$ . The set of mixed strategies of  $i$  is denoted by  $\Delta(S_i)$ . The payoff function for  $i$  is  $h_i: S \rightarrow R$ . It is extended in the usual way to  $\times_i \Delta(S_i)$ . In order to describe iterated elimination of dominated strategies we use the following terminology.

**Definition 1.** (relative domination) Let  $T_{-i}$  be a nonempty set of profiles of  $i$ 's opponents. A mixed strategy  $\sigma_i \in S_i$  *strongly dominates*  $s_i$  *relative to*  $T_{-i}$  if  $h_i(\sigma_i, t_{-i}) > h_i(s_i, t_{-i})$  for all  $t_{-i} \in T_{-i}$ . We say in this case that  $s_i$  is *strongly dominated relative to*  $T_{-i}$ . The strategy  $\sigma_i$  *weakly dominates*  $s_i$  *relative to*  $T_{-i}$  if  $h_i(\sigma_i, t_{-i}) \geq h_i(s_i, t_{-i})$  for all  $t_{-i} \in T_{-i}$ , and at least one of these inequalities is strict. We say in this case that  $s_i$  is *weakly dominated relative to*  $T_{-i}$ .

Using this terminology we define the processes of elimination of dominated strategies.

**Definition 2.** A process of *iterated elimination of strongly dominated strategies* consists of sequences of strategy profile sets  $(S^0, S^1, \dots, S^m)$ , where  $S^0 = S$ , and for  $k \geq 1$ ,  $S^k = \times_i S_i^k$ , where  $S_i^k$  is obtained from  $S_i^{k-1}$  by eliminating some strategies in the latter set which are strongly dominated relative to  $S_{-i}^{k-1}$ . In the sets  $S_i^m$  there are no strongly dominated strategies relative to  $S_{-i}^m$ . A process of *iterated elimination of weakly dominated strategies* is similarly defined, where in each stage, weakly dominated strategies are eliminated.<sup>2</sup>

<sup>2</sup>A strategy is strongly dominated by a mixed strategy if and only if it is not a best response against any probability distribution on the opponents profiles. Thus, it is possible to describe the iterated elimination of strongly dominated strategies as the iterative elimination of strategies that are not a best response (see Myerson, 1991, 88–89). Bernheim (1984) and Pearce (1984) suggested an iterative process that terminates with *rationalizable* profiles. In their process, strategies are eliminated when they are not a best response against probabilistic distribution over the opponents profiles, where player's strategies are drawn independently.

There can be several processes of iterative elimination of strongly dominated strategies. However, it is well known that all such processes end in the same set of strategy profiles, to which we refer as the set of profiles that survive iterated elimination of strongly dominated strategies. In contrast, different processes of iterative elimination of weakly dominated strategies may end in different sets of strategy profiles.

The property of relative strong dominance that explains why all processes of iterative elimination of strongly dominated strategy terminate in the same set is the following monotonicity property, which is straightforward to show.

**Claim 1.** (strong-dominance monotonicity) *If a strategy of  $i$  is strongly dominated relative to  $T_{-i} \subseteq S_{-i}$  then it is also strongly dominated relative to  $T'_{-i} \subseteq T_{-i}$ .*

Relative weak dominance does not have this monotonicity property.

### 3. STRONG DOMINANCE

**3.1. Informal justification of the iterated process.** The process of iterated elimination of strongly dominated strategies, described above, can be justified by assuming common knowledge of rationality. We consider this justification, which is described below, informal, since the term knowledge is not formalized.

Rationality, here, means, or implies, that players do not play strongly dominated strategies. Common knowledge of rationality is tantamount to saying that all players are rational, they all know it, they all know that they all know it and so on. We can now justify the process of the strategy elimination in steps.

- (1) All players are rational, and thus the strategy profile they play must be in  $S^1$ .
- (2) Moreover, all players know that all players are rational, and thus they all know that the profile played is in  $S^1$ . Being rational, the strategy profile they play must be in  $S^2$ .
- (3) Moreover, all players know that all players know that all players are rational, and thus ... and so on.

Note, that in step  $k$  each player  $i$  knows that her opponents are playing a profile in  $S_{-i}^{k-1}$ , and this is why dominance is considered only relative to this set. According to this explanation of the process, a player is rational if she does not play a strongly dominated strategy relative to the set of her opponents' profiles that her knowledge does not exclude. We call this kind of rationality, *strong-dominance rationality*, and define it formally later.

This informal argument for justifying the iterative elimination of strongly dominated strategies raises the following problem. By this argument, common knowledge of rationality implies that each player  $i$  knows that her opponents are playing a profile in  $S_{-i}^m$ . But, in the process we eliminate strongly dominated strategies given that player  $i$  knows less than that. That is, we assume along the way that player  $i$  knows that the other players play  $S_{-i}^0$  and then  $S_{-i}^1$  and so on. This seeming contradiction is spurious because of the monotonicity of relative strong dominance in Claim 1: if a strategy is eliminated in stage  $k$  when player  $i$  knows that her opponents' profile is in  $S_{-i}^{k-1}$ , then it should be eliminated also when she knows more than that, that is, when she knows that her opponents' profile is in  $S_{-i}^m$ .

**3.2. The formal justification of the iterated process.** We formalize the argument of the previous subsection by using a state space in which knowledge is formally defined. Such a model enables us to directly capture common knowledge of rationality as an event in the state space, without the hierarchy of knowledge. We show that this modeling reaffirms the justification of the iterated process.<sup>3</sup>

Let  $\Omega$  be a finite state space with a partition  $\Pi_i$  for each player  $i$ . At a state  $\omega$  player  $i$  knows all the events that contain  $\Pi_i(\omega)$ , the element of  $i$ 's partition that contains  $\omega$ . For simplicity we assume that the meet of the partition consists of  $\Omega$ . Thus, in each state,  $\Omega$  is the only event that is common knowledge (see Aumann, 1976). In order to define knowledge about strategy profiles, we assume that each state determines the strategy profile played in the state. The strategies played in each state are given by functions  $\mathbf{s}_i: \Omega \rightarrow S_i$ , such that  $\mathbf{s}_i(\omega)$  is the strategy  $i$  plays at state  $\omega$ . We further assume that each player knows which strategy she plays. This means that  $\mathbf{s}_i$  is measurable with respect to  $\Pi_i$ , or in other words, in each element of  $\Pi_i$ ,  $i$  plays the same strategy in all the states in the element.

We can now define the event that a player is rational, in agreement with the informal definition of the previous subsection.

**Definition 3.** Player  $i$  is *strong-dominance rational* in state  $\omega$  if the strategy she plays in  $\omega$  is not strongly dominated relative to the set of her opponents' profiles which are not excluded by her knowledge in  $\omega$ . That is, there is no strategy of hers that strongly dominates  $\mathbf{s}_i(\omega)$  relative to the set  $T_{-i}(\omega) = \{\mathbf{s}_{-i}(\omega') \mid \omega' \in \Pi_i(\omega)\}$ .

The antecedent in the following proposition is that it is common knowledge that all players are strong-dominance rational. By our assumption that  $\Omega$  is the only element of the meet, this is equivalent to saying that each of the players is strong-dominance rational in each state.

**Proposition 1.** *If it is common knowledge that the players are strong-dominance rational, then the strategy profiles played survive the iterated elimination of strongly dominated strategies.*

*Proof.* The proof, not surprisingly, resembles the iterated elimination of strongly dominated strategies, using the monotonicity property in Claim 1. Since strong-dominance rationality of all players is common knowledge, players are strong-dominance rational in each state. Therefore, for any state  $\omega$  and player  $i$ ,  $\mathbf{s}_i(\omega)$  must be a strategy in  $S_i^1$ . Otherwise, it is strongly dominated in the game and therefore by monotonicity and the strong-dominance rationality of  $i$ , when  $i$  knows  $T_{-i}(\omega)$  she cannot play it in  $\omega$ . Thus, in all states the profiles played are in  $S^1$ . But then, for all  $\omega$  and  $i$ ,  $T_{-i}(\omega) \subseteq S_{-i}^1$ , and by the same argument as above, all the profiles played must be in  $S^2$ , and so on.  $\square$

#### 4. WEAK DOMINANCE

The informal argument for iterated elimination of strongly dominated strategies can be stated verbatim for the case of weakly dominated strategies by interpreting rationality as *weak-dominance rationality*. That is, a player is weak-dominance rational if she does not play a weakly dominated strategy relative to the set of her

<sup>3</sup>As mentioned in footnote 1, eliminating dominated strategies is the same as eliminating strategies that are not a best response. Tan and Werlang (1988) formalize the iterative elimination of non-best response strategies in probabilistic belief spaces. Their characterization can be considered as dual to ours.

	L	C	R		L	C	R		L	C	R		L	C	R
T	1, 0	2, 0	3, 0		1, 0	2, 0	3, 0		1, 0	2, 0	3, 0		1, 0	2, 0	3, 0
M	1, 2	3, 1	2, 1		1, 2	3, 1	2, 1		1, 2	3, 1	2, 1		1, 2		
B	1, 1	0, 3	4, 2		1, 1	0, 3			1, 1				1, 1		

FIGURE 1. Common knowledge of weak-dominance rationality

opponent's profiles that she does not exclude. However, the argument that common knowledge of rationality implies iterative elimination of weakly dominated strategies is flawed. Suppose that a process of iterative elimination of weakly dominated strategies ends with a set of profiles  $S^m$ . If, as the argument goes, common knowledge of weak-dominance rationality implies that the strategy profile must be in  $S^m$ , then players know it. Thus, weak-dominance rationality requires only that each player  $i$  does not play a strategy which is weakly dominated relative to  $S_{-i}^m$ . However, along the process we eliminated strategies of  $i$  relative to sets of profiles that reflect less knowledge. Since weak dominance does not have the monotonicity property in Claim 1, these eliminated strategies may not be weakly dominated relative to  $S_{-i}^m$  and thus their elimination is not justified in view of the player's knowledge.

In the following example we demonstrate, at this point still informally, how common knowledge of weak-dominance rationality implies a process of elimination different from iterative elimination of weakly dominated strategies.

**Example 1.** Consider the game on the left side of Figure 1, and assume that it is common knowledge that the players are weak-dominance rational. There is a unique dominated strategy in this game, strategy R, which is weakly dominated by C. However, eliminating strategy R may be unjustified, as was pointed out in explaining why the informal argument for iterative elimination of weakly dominated strategies is flawed; It is possible that player 2 *does* play R if she knows that player 1 does not play B. Since we have not discovered at this stage what player 2 knows about player 1's strategy, we cannot eliminate the possibility that she plays R. But, obviously, if player 2 is weak-dominance rational and she *does* play R it must be because she knows that player 1 *does not* play B. Thus, the only thing we can infer from player 2's weak-dominance rationality is that the profile (B,R) cannot possibly be played. We refer to this profile as a *dominance base* for the weakly dominated strategy R. We eliminate this profile and get the set of profiles with non-blank payoffs in the second table from the left, which we denote by  $T^1$ . We conclude that when players are weak-dominance rational, the profile played should be from this set.

Now, since the players know that they are weak-dominance rational, they know that the profile played is in  $T^1$ . Thus, if player 1 plays B, she excludes the possibility that player 2 plays R, and she knows that player 2 plays either L or C. Relative to this knowledge, strategy B is weakly dominated by M. However, we cannot eliminate strategy B since it is possible that player 1 knows that player 2 plays L in which case B is not weakly dominated by M. But we can conclude by player 1's weak-dominance rationality, that if she does play B it must be the case that she knows that the profile (B,C), the dominance base for B relative to player 1's knowledge, is not played. Hence, we eliminate this profile and the set  $T^2$  of the

remaining profiles is depicted in the third table from the left. We conclude that when players know that they are weak-dominance rational then the profile they play must be in  $T^2$ .

When players know that they know that they are weak-dominance rational, then they know that the profile played is in  $T^2$ . If player 1 plays either C or R then, being weak-dominance rational, it must be the case that she excludes the possibility that player 1's strategy is M, because in that case L is weakly dominant, given her knowledge, which contradicts her weak-dominance rationality. Thus, we eliminate (M,C) and (M,R), the dominance bases of M, and the remaining set of profiles,  $T^3$ , is presented in the last table.

At this stage we cannot eliminate any profile from  $T^3$  and we conclude that if there is common knowledge of weak-dominance rationality, then the strategy profile played should be in  $T^3$ .

Note that there exists a unique process of iterated elimination of weakly dominated strategies, in which the order of elimination is R, B, C. The remaining profiles are (T,L) and (M,L). Thus this process results in a smaller set of profiles than the process described above. However, common knowledge of weak-dominance rationality does not imply that only one of these two profiles is played. Indeed, if this were the case, then under the assumption of common knowledge of weak-dominance rationality player 1 knows that player 2 is playing L. But, then B is not weakly dominated given player 1's knowledge. The elimination of B is justified only if C is considered possible by player 1 when she plays B.

**4.1. Eliminating dominance bases.** We generalize the process described in Example 1; it is not the weakly dominated strategy which is the culprit that has to be eliminated, but rather the profile in which this strategy is played in which the player is strictly worse off in comparison to the dominating strategy. We define it formally.

**Definition 4.** The profile  $(s_i, t_{-i})$  is a *weak-dominance base* for  $s_i$  relative to  $T_{-i} \subseteq S_{-i}$ , if  $s_i$  is weakly dominated relative to  $T_{-i}$ ,  $t_{-i}$  is in  $T_{-i}$ , and  $h_i(\sigma_i, t_{-i}) > h_i(s_i, t_{-i})$  for some mixed strategy  $\sigma_i$  of  $i$  that weakly dominates  $s_i$  relative to  $T_{-i}$ . A *strong-dominance base* for  $s_i$  relative to  $T_{-i}$  is similarly defined by changing “weakly” to “strongly”.

It is straightforward to show that both weak and strong dominance bases have the same monotonicity property that strongly dominated strategies have.

**Claim 2.** (monotonicity of dominance bases) *If  $(s_i, t_{-i})$  is a weak-dominance (strong-dominance) base for  $s_i$  relative to  $T_{-i}$  then it is also a weak-dominance (strong-dominance) base for  $s_i$  relative to  $T'_{-i} \subseteq T_{-i}$ .*

A process of *iterative elimination of weak-dominance bases* consists of sequence of strategy profile sets  $S^0, S^1, \dots, S^m$ , where  $S^0 = S$ , and for each  $k$ ,  $S^k$  is obtained by eliminating from  $S^{k-1}$  some profiles  $(s_i, t_{-i})$  for some players  $i$ , such that  $(s_i, t_{-i})$  is a weak-dominance base for  $s_i$  relative to  $\{t'_{-i} \mid (s_i, t'_{-i}) \in S^{k-1}\}$ . A process of *iterative elimination of strong-dominance bases* is defined *mutatis mutandis*.

Due to the monotonicity property in Claim 2, iterative elimination of dominance bases has the desired property that all processes end in the same set of profiles.

**Proposition 2.** *All processes of iterative elimination of weak-dominance bases end at the same set of profiles, and all processes of iterative elimination of strong-dominance bases end at the same set of profiles.*

*Proof.* Suppose that  $\hat{S}$  is the last set of profiles in a process of iterative elimination of weak-dominance bases. Consider another such process:  $S^0, S^1, \dots, S^m$ . We show by induction on  $k$  that  $\hat{S} \subseteq S^k$ . Thus, in particular,  $\hat{S} \subseteq S^m$  which shows that any two such processes end at the same set. As  $S^0 = S$ , the claim for  $k = 0$  is obvious. Suppose that  $\hat{S} \subseteq S^k$ , for  $k < m$ . If profile  $(s_i, t_{-i})$  was eliminated from  $S^k$ , then it is a dominance base of  $s_i$  relative to  $\{t'_{-i} \mid (s_i, t'_{-i}) \in S^k\}$ . Suppose that this profile is in  $\hat{S}$ . As  $\hat{S} \subseteq S^k$ ,  $\{t'_{-i} \mid (s_i, t'_{-i}) \in \hat{S}\} \subseteq \{t'_{-i} \mid (s_i, t'_{-i}) \in S^k\}$ , and thus by Claim 2,  $(s_i, t_{-i})$  is a dominance base of  $s_i$  relative to  $\{t'_{-i} \mid (s_i, t'_{-i}) \in \hat{S}\}$  contrary to the assumption that no profiles can be eliminated from  $\hat{S}$ . We conclude that all the profiles that are eliminated from  $S^k$  are not elements of  $\hat{S}$ , and therefore,  $\hat{S} \subseteq S^{k+1}$ . The proof for the case of strong-dominance bases is the same.  $\square$

Note, that if  $s_i$  is strongly dominated relative to  $T_{-i}$ , then for *all*  $t_{-i} \in T_{-i}$ ,  $(s_i, t_{-i})$  is a strong-dominance base for  $s_i$  relative to  $T_{-i}$ . Thus, a process of elimination of strongly dominated strategies is in particular a process of elimination of strong-dominance bases. Thus, in view of Proposition 2 we conclude:

**Corollary 1.** *The set of profiles that survive iterative elimination of strongly dominated strategies coincides with the set of profiles that survive iterative elimination of strong-dominance bases.*

However, as demonstrated in Example 1, the set of profiles that survive iterative elimination of weak-dominance bases may differ from any set of profiles that end a process of iterative elimination of weakly dominated strategies. It is the first process that captures common knowledge of weak-dominance rationality, as we show it in the next subsection.

**4.2. Common knowledge of weak dominance rationality formalized.** We first define weak-dominance rationality analogously to strong-dominance rationality in Definition 3, and then state our main result.

**Definition 5.** Player  $i$  is *weak-dominance rational* in state  $\omega$  if the strategy she plays at  $\omega$  is not weakly dominated relative to the set of her opponents' profiles which are not excluded by her knowledge at  $\omega$ . That is, there is no strategy of hers that weakly dominates  $\mathbf{s}_i(\omega)$  relative to the set  $T_{-i}(\omega) = \{\mathbf{s}_{-i}(\omega') \mid \omega' \in \Pi_i(\omega)\}$ .

**Proposition 3.** *If it is common knowledge that the players are weak-dominance rational, then the strategy profiles played survive the iterative elimination of weak-dominance bases.*

*Proof.* The proof, like that of Proposition 1, mimics the iterative process. Let  $S^0, \dots, S^m$  be a process of iterative elimination of weak dominance bases. Since weak-dominance rationality of all players is common knowledge, players are weak-dominance rational in each state. Therefore, for any state  $\omega$  and player  $i$ ,  $\mathbf{s}(\omega)$  must be a strategy in  $S^1$ . Otherwise, for some  $\omega$  and  $i$ ,  $\mathbf{s}(\omega) = (s_i, s_{-i})$  where  $s_i$  is weakly dominated relative to  $S_{-i}$ , and  $(s_i, s_{-i})$  is a weak-dominance base for  $s_i$  relative to  $S_{-i}$ . But, if so, then by Claim 2,  $s_i$  is also a weak-dominance base for  $s_i$  relative to  $T_{-i}(\omega) = \{\mathbf{s}_{-i}(\omega') \mid \omega' \in \Pi_i(\omega)\}$ . This implies that  $s_i$  is weakly dominated with respect to this set, which contradicts our assumption.

Now, as  $\mathbf{s}(\omega) \in S^1$  for each  $\omega$ ,  $T_{-i}(\omega) = \{\mathbf{s}_{-i}(\omega') \mid \omega' \in \Pi_i(\omega)\} \subseteq \{s'_{-i} \mid (s_i, s'_{-i}) \in S^1\}$ . Therefore, for each  $\omega$ ,  $\mathbf{s}(\omega) \in S^2$ , or else, for some  $i$ ,  $\mathbf{s}(\omega) =$

$(s_i, s_{-i})$  is a weak-dominance base for  $s_i$  relative to  $\{s'_{-i} \mid (s_i, s'_{-i}) \in S^1\}$  and hence, by monotonicity, also relative to  $T_{-i}(\omega)$ , which contradicts the weak-dominance rationality of  $i$  at  $\omega$ . The argument for the next stages is similar.  $\square$

In light of Corollary 1, Proposition 1 can be restated analogously to Proposition 3.

**Proposition 4.** *If it is common knowledge that the players are strong-dominance rational, then the strategy profiles played survive the iterative elimination of strong-dominance bases.*

We conclude that iterative elimination of dominance bases captures both the case of common knowledge of strong-dominance rationality and common knowledge of weak dominance rationality. Iterative elimination of dominated *strategies* in the strong dominance case coincides with iterative elimination of strong-dominance bases. This coincidence does not hold in the weak dominance case, and the iterative elimination of weakly dominated strategies fails to capture common knowledge of weak-dominance rationality.

#### REFERENCES

- ROBERT J. AUMANN (1976): “Agreeing to disagree,” *Ann. Statist.*, 4(6), 1236–1239.
- B. DOUGLAS BERNHEIM (1984): “Rationalizable strategic behavior,” *Econometrica*, 52(4), 1007–1028.
- KENNETH BINMORE (1991): *Fun and Games: A Text on Game Theory*, D. C. Heath and Company.
- JOHN HILLAS AND DOV SAMET (2013): “Non-probabilistic correlated equilibrium as an expression of non-Bayesian rationality,” a manuscript.
- ROGER MYERSON (1991): *Game Theory: Analysis of Conflict*, Harvard University Press, Cambridge, MA.
- DAVID G. PEARCE (1984): “Rationalizable strategic behavior and the problem of perfection,” *Econometrica*, 52(4), 1029–1050.
- TOMMY CHIN-CHIU TAN AND SÉRGIO RIBEIRO DA COSTA WERLANG (1988): “The Bayesian foundations of solution concepts of games,” *Journal of Economic Theory*, 45(2), 370–391.



# NON-PROBABILISTIC CORRELATED EQUILIBRIUM AS AN EXPRESSION OF NON-BAYESIAN RATIONALITY

JOHN HILLAS AND DOV SAMET

ABSTRACT. We define *weak* and *strong rationality* of players in terms of dominance rather than expectation with respect to probabilistic beliefs. We also define certain subsets of strategy profiles as *weak* or *strong non-probabilistic correlated equilibrium*, analogously to correlated equilibrium. It is shown that the set of profiles played when there is common knowledge of weak or strong rationality is a weak or strong non-probabilistic correlated equilibrium, correspondingly. The largest weak non-probabilistic correlated equilibrium is the set of profiles that survive iterated elimination of strictly dominated strategies. Rationality can be strengthened for games with perfect information by considering rationality in subgames. A player is *substantively rational* if she is rational in all subgames, and *materially rational* if she is rational in subgames at vertices that are reached. Aumann (1995, 1998) introduced and studied these two notions for the case of weak rationality. We characterize the weak and the strong versions of these notions of rationality in terms of non-probabilistic correlated equilibria. We show that non-probabilistic correlated equilibria can be also characterized in terms of non-probabilistic belief rather than knowledge.

## 1. INTRODUCTION

Rationality in economic theory and game theory is most commonly assumed to be Bayesian. That is, each player has beliefs about the strategies adopted by her opponents, and her strategy maximizes her utility with respect to these beliefs. This is true in particular in most works in which the implications of rationality are explicitly addressed. Bayesian rationality is the main building block of rationalizability as defined by Bernheim (1984) and Pearce (1984), and in the studies of its variants in Tan and Werlang (1988) and Brandenburger and Dekel (1987). In a paper titled “Correlated equilibrium as an expression of Bayesian rationality,” Aumann (1987) showed that common knowledge of players’ Bayesian rationality, when players’ beliefs are derived from a common prior, implies that the prior probability of the strategy profiles played is a correlated equilibrium, a notion that he introduced in 1974. We note that probabilistic beliefs are used even in cases where they characterize solutions that are defined without resorting to probabilities of opponents’ strategies. Tan and Werlang (1988) and Brandenburger and Dekel (1987) characterize the set of strategies that survive iterated elimination of strictly dominated strategies in terms of probabilistic beliefs although this set is defined without using such beliefs. Similarly, Brandenburger et al. (2008) use lexicographic probabilistic beliefs to characterize a set of admissible strategies, where again, the set itself is defined without resorting to probabilistic beliefs.

Here we study notions of rationality without making the very basic assumption of Bayesianism that players form probabilistic beliefs about other players' strategies. Two notions of rationality defined in non-probabilistic terms were already studied in Aumann (1995, 1998) for the analysis of games in extensive form. Our analysis starts with games in strategic form, characterizes the solutions defined by common knowledge of rationality, and shows how the theory for games in extensive form in Aumann (1995, 1998) can be expressed in terms of these solutions and even reduced to them.

We say that a player is *weakly rational* if there is no strategy of hers that she knows yields her a higher payoff than the one she is playing. Equivalently, the player is weakly rational if there is no strategy of hers that yields her a payoff higher than yielded by her actual strategy, against any of her opponents' strategy profiles that she considers possible. A player is *strongly rational* if there is no strategy of hers that she knows yields her at least as high payoff as yielded by the strategy she is playing, unless she knows that this strategy yields her the same payoff as the one she is playing. That is, there is no strategy of hers that yields her at least as high payoff as her actual strategy against any of her opponents' strategy profiles that she considers possible, and a higher payoff at least against one such profile.

Strong rationality obviously implies weak rationality. In models of knowledge and probabilistic belief both can be compared to Bayesian rationality. Bayesian rationality implies weak rationality and may be, and typically is, strictly stronger. But, in such models, strong rationality is incomparable with Bayesian rationality. Clearly, a player can be strongly rational but not Bayesian rational. Conversely, a player can be Bayesian rational, but fail to be strongly rational if there is a strategy of hers that is equivalent to her Bayesian optimal strategy with probability one, but dominates the Bayesian optimal strategy when states of probability zero are considered. However, when strong rationality is defined in terms of belief with probability 1, rather than knowledge, this cannot happen, and in this case Bayesian rationality implies strong rationality. We discuss weak and strong rationality in terms of belief later in this section and in Section 6.

Non-probabilistic correlated equilibria are defined similarly to probabilistic ones. We think of a mediator who is choosing any strategy profile  $s$  from a given set of strategy profiles  $C$ . She then suggests to each Player  $i$  to play  $s_i$ . The set  $C$  is a *weak non-probabilistic correlated equilibrium* if there is no strategy of Player  $i$  that yields a higher payoff than yielded by  $s_i$  against each of the opponents' strategy profiles that can be suggested to them when the suggestion to  $i$  is to play  $s_i$ . The set  $C$  is a *strong non-probabilistic correlated equilibrium* if there is no strategy of Player  $i$  that yields at least as high a payoff as yielded by  $s_i$  against each of the opponents' strategy profiles that can be suggested to them when the suggestion to  $i$  is to play  $s_i$ , and a strictly higher payoff against at least one of these strategy profiles.

We have called the concept we defined a non-probabilistic correlated equilibrium, in spite of the fact that it does not involve correlated probabilities and it is not a Nash equilibrium of the game with the mediator. However, non-probabilistic correlated equilibria have properties that are analogous to probabilistic correlation and Nash equilibrium. A probability distribution over the set of strategy profiles is said to be correlated when it is not the independent product of probability distributions on the strategy sets of the individual players. Similarly, we may call a

set of profiles correlated when it is not the product of subsets of the strategy sets of the individual players. Although the behavior of the players in the game with the mediator is not a Nash equilibrium, it is in the spirit of equilibrium behavior, since players maximize in the following sense. A player's choice is a strategy which is maximal in the partial order of strategies defined by domination relative to her knowledge. Note, also, that the notion of non-probabilistic correlated equilibrium extends the notion of Nash equilibrium: The support of the distribution over strategy profiles induced by any Nash equilibrium, and hence any singleton consisting of a pure strategy Nash equilibrium, is obviously a strong non-probabilistic correlated equilibrium, and hence also a weak non-probabilistic correlated equilibrium.

We use a standard model of knowledge in order to translate simple statements in natural language about rationality, knowledge, and common knowledge into subsets (events) in the model. This is the approach used in Aumann (1987), but without the extra feature that formalizes probabilistic beliefs. We can now state the characterization of non-probabilistic correlated equilibria as an expression of non-Bayesian rationality.

*A non-empty set of strategy profiles is a weak (strong) non-probabilistic correlated equilibrium, if and only if it is the set of strategy profiles played when weak (strong) rationality is common knowledge.*

The families of weak and strong non-probabilistic correlated equilibria are not empty and have each a largest element. We describe a process of iterated elimination of profiles that results in the largest non-probabilistic correlated equilibrium. The largest weak non-probabilistic correlated equilibrium is the set of profiles that survive iterated elimination of strongly dominated strategies.<sup>1</sup> This set has been previously characterized in terms of Bayesian rationality. For two player games, this is the set of rationalizable strategy profiles of Bernheim (1984) and Pearce (1984). For more than two players, it was characterized by Tan and Werlang (1988) and by Brandenburger and Dekel (1987). Our result provides a characterization in terms of non-Bayesian weak rationality, which is weaker than Bayesian rationality. We also characterize all the *sets* of profiles that may arise when weak rationality is common knowledge.

Strong rationality is defined in terms of weak dominance, relative to a player's knowledge, and not surprisingly it is related to elimination of weakly dominated strategies. Indeed, every set of profiles that survive iterated elimination of weakly dominated strategies is a strong non-probabilistic correlated equilibrium. However a strong non-probabilistic correlated equilibrium may contain other profiles as well. For example, every singleton that consists of a pure strategy equilibrium profile is a strong non-probabilistic correlated equilibrium, even when in this equilibrium some of the players' strategies are weakly dominated.

The centipede game can serve to demonstrate the difference between various solutions. Since there are no strongly dominated strategies in this game, the largest weak non-probabilistic correlated equilibrium is the set of all profiles. Iterated elimination of weakly dominated strategies results in a unique profile, that of the unique pure strategy equilibrium. The largest strong non-probabilistic correlated

---

<sup>1</sup>In the literature the notion we are calling strong dominance is referred to both as strict dominance and as strong dominance, most commonly as strict dominance. Here, for the coherency of terminology, we call it strong dominance.

equilibrium lies between these two solutions. It contains all strategy profiles in which the first player terminates the game immediately.

For games in extensive form of perfect information, it is possible to require that a player is rational not only in the whole game, but also in some or all of its subgames. Aumann (1995) applied to games in general position the requirement of *substantive rationality*, which in our terminology means that a player is weakly rational in all the subgames. Aumann (1998) applied to the centipede game the requirement of *material rationality*, which requires that a player is weakly rational only in subgames at reached vertices. We explore here all four possible notions of rationality with the attributes weak-strong and substantive-material, and define the corresponding non-probabilistic correlated equilibria characterized by common knowledge of these types of rationality. Indeed, there are only three notions of rationality, as we show that material strong rationality is the same as strong rationality. Moreover, for games in general position, even material weak rationality is the same as strong rationality. Thus, the result of Aumann (1998) which shows that common knowledge of material weak rationality implies in the centipede game that the first player terminates the game immediately, follows from the fact that this holds for the strong non-probabilistic correlated equilibria of this game.

So far we defined rationality in terms of knowledge. The assumption that a player has some knowledge about her opponents' strategies, let alone their rationality, and thus never errs, is a strong assumption. This is why we study rationality in terms of (non-probabilistic) belief which has all the properties of knowledge, except that, unlike knowledge, what is believed is not necessarily true. We show that the difference between knowledge and belief is not crucial for the theory presented here, which justifies the use of the simpler model of knowledge, despite its strength. We describe as *epistemic*, matters of knowledge, and as *doxastic*, matters of belief.

Doxastic weak rationality is defined similarly to epistemic weak rationality with belief replacing knowledge. Here there is no difference between knowledge and belief, since weak non-probabilistic correlated equilibrium is characterized by common belief of doxastic weak rationality in precisely the same way it is characterized by common knowledge of epistemic weak rationality.

For the characterization of strong non-probabilistic correlated equilibrium by doxastic strong rationality a consistency condition is required in much the same way that a similar condition is required for the characterization of correlated equilibrium in Aumann (1987). There, consistency means that players' beliefs are derived from a common prior.<sup>2</sup> Here we require *interpersonal consistency of non-probabilistic beliefs*, which was used by Bonanno and Nehring (1998) and Samet (2013) to derive game theoretical results. It says that each player not only believes that her beliefs are correct (which follows from the properties of belief) but also believes that her opponents' beliefs are correct. The notion of consistency of belief can be compared to consistency of probabilistic beliefs in models of knowledge and probabilistic belief. In such models, the existence of a common prior implies that belief with probability 1 is consistent in the sense we defined here. Knowledge by this definition is interpersonally consistent by virtue of being true.

---

<sup>2</sup>Although this seems to require a dynamic process of updating, consistency can be equivalently defined in static terms as the impossibility of some absurd bets (Feinberg, 2000; Morris, 1994; Samet, 1998a) or the convergence of all possible iterated expectations to the same result (Samet, 1998b). The notion of interpersonal consistency of non-probabilistic beliefs defined here is straightforwardly static.

We show that a set of profiles is a strong non-probabilistic correlated equilibrium if and only if it is the set of profiles which are played when there is common belief of doxastic strong rationality where players beliefs are interpersonally consistent.

## 2. PRELIMINARIES

**2.1. Games in strategic form.** Let  $G$  be a game in strategic form with a finite set of players  $I$ . The set of strategies for each  $i \in I$  is  $S_i$ . We denote  $S = \times_{i \in I} S_i$  and  $S_{-i} = \times_{j \in I \setminus \{i\}} S_j$ . The elements of  $S$  are called *strategy profiles* or *profiles* for short. The elements of  $S_{-i}$  are the *profiles of  $i$ 's opponents*.

We let  $s_i$  denote both an element of  $S_i$ , and the projection on  $S_i$  of a profile  $s \in S$ . Similarly, we let  $s_{-i}$  denote both an element of  $S_{-i}$  and the projection on  $S_{-i}$  of such a profile  $s \in S$ . For any finite set  $X$ ,  $\Delta(X)$  is the set of probability distributions over  $X$ . The set  $\Delta(S_i)$  consists of the mixed strategy of Player  $i$ . Player  $i$ 's payoff function is  $h_i: S \rightarrow R$ . We extend the definition of  $h_i$  to profiles of mixed strategies in the usual way by taking expectations.

**2.2. Knowledge structures.** A *knowledge structure for the game  $G$*  is a tuple  $(\Omega, (\Pi_i)_{i \in I}, \mathbf{s})$  which consists of a finite set of states  $\Omega$ , a partition  $\Pi_i$  of  $\Omega$  for each  $i$ , and a function  $\mathbf{s}: \Omega \rightarrow S$  that describes the strategy profile at each state. We assume that for each  $i$ ,  $\mathbf{s}_i$  is constant on each element of  $\Pi_i$ , which means that  $i$  always knows her strategy. Subsets of  $\Omega$  are called *events*. The knowledge operator  $K_i$ , that maps each event  $E$  to the event  $K_i E$  that reads “ $i$  knows  $E$ ”, is defined by  $K_i E = \{\omega \mid \Pi_i(\omega) \subseteq E\}$ , where  $\Pi_i(\omega)$  is the element of  $\Pi_i$  that contains  $\omega$ . The event that *all know  $E$*  is  $KE = \cap_{i \in I} K_i E$ . The event  $CKE$ , that  $E$  is *common knowledge*, is the event that all know  $E$ , all know that all know  $E$  and so on. Thus,  $CKE = \cap_{n \geq 1} K^n E$ , where  $K^n$  is the  $n$ th power of  $K$ . The event  $CKE$  is a union of elements of  $\Pi_i$  for each  $i$ , and it is the largest event of this kind that is a subset of  $E$ .

**2.3. Games in extensive form with perfect information.** For a game in extensive form with perfect information, we denote by  $V$  the set of nonterminal vertices, which we refer to in the sequel as vertices. For each  $v \in V$ , we denote by  $G^v$  the strategic form of the game that starts in vertex  $v$ . We write  $G$  for the whole game, that starts at the root. The payoff functions in  $G^v$  are denoted by  $h_i^v$ , the strategy sets are  $S_i^v$ ,  $S_{-i}^v$  and  $S^v$ . For any subset of strategies or profiles  $X$  in the game  $G$  we write  $X^v$  for the restriction of profiles in  $X$  to vertices in the subtree that starts at  $v$ . If  $i$  does not have vertices in this subtree, then  $S_i^v$  is a singleton. Note that  $S_i^v$  denotes both the set of  $i$ 's strategies in  $G^v$  and the restriction of  $S_i$  to the subtree at  $v$ . No confusion results, as the two sets coincide. We can view  $h_i^v$  as a function defined on  $S$ , by setting  $h_i^v(s) = h_i^v(s^v)$ . A knowledge structure for  $G$ ,  $(\Omega, (\Pi_i)_{i \in I}, \mathbf{s})$  induces for each  $v$  a knowledge structure  $(\Omega, (\Pi_i)_{i \in I}, \mathbf{s}^v)$  for  $G^v$ , where  $\mathbf{s}^v(\omega) = (\mathbf{s}(\omega))^v$ . Thus, we can view  $(\Omega, (\Pi_i)_{i \in I}, \mathbf{s})$  as being a knowledge structure for all the games  $G^v$  simultaneously. Every profile  $s \in S$  determines a path in the game tree. We say that  $s$  *reaches*  $v \in V$  if  $v$  is on this path. A game is in *general position* if the payoffs to each player at different leaves of the game tree are different.

### 3. COMMON KNOWLEDGE OF RATIONALITY

We introduce two notions of rationality for games in strategic form, and two solutions for such games that are characterized by common knowledge of rationality. We first define the notion of relative domination that is used in the sequel.

**Definition 1.** Let  $T_{-i} \subseteq S_{-i}$  be a non-empty set of strategy profiles of  $i$ 's opponents.

We say that a strategy  $s_i \in S_i$  is *strongly dominated relative* to  $T_{-i}$ , if there is some  $\sigma_i \in \Delta(S_i)$  such that  $h_i(\sigma_i, t_{-i}) > h_i(s_i, t_{-i})$  for each  $t_{-i} \in T_{-i}$ , in which case we say that  $s_i$  is strongly dominated by  $\sigma_i$ , relative to  $T_{-i}$ .

We say that a strategy  $s_i \in S_i$  is *weakly dominated relative* to  $T_{-i}$ , if there is some  $\sigma_i \in \Delta(S_i)$  such that  $h_i(\sigma_i, t_{-i}) \geq h_i(s_i, t_{-i})$  for each  $t_{-i} \in T_{-i}$ , where at least one of the inequalities is strict, in which case we say that  $s_i$  is weakly dominated by  $\sigma_i$ , relative to  $T_{-i}$ .

**3.1. Rationality.** We first define rationality of a player in a given state, and then describe the event that a player is rational in terms of the player's knowledge.

**Definition 2.** Player  $i$  is *weakly rational* at state  $\omega$  if  $\mathbf{s}_i(\omega)$  is not strongly dominated relative to  $\{\mathbf{s}_{-i}(\omega') \mid \omega' \in \Pi_i(\omega)\}$ , that is, if there is no strategy  $\sigma_i \in \Delta(S_i)$  such that  $h_i((\sigma_i, \mathbf{s}_{-i}(\omega'))) > h_i(\mathbf{s}(\omega'))$  for all  $\omega' \in \Pi_i(\omega)$ .

Player  $i$  is *strongly rational* at state  $\omega$ , if  $\mathbf{s}_i(\omega)$  is not weakly dominated relative to  $\{\mathbf{s}_{-i}(\omega') \mid \omega' \in \Pi_i(\omega)\}$ , that is, if there is no strategy  $\sigma_i \in \Delta(S_i)$  such that  $h_i((\sigma_i, \mathbf{s}_{-i}(\omega'))) \geq h_i(\mathbf{s}(\omega'))$  for all  $\omega' \in \Pi_i(\omega)$ , with at least one of these inequalities being strict.

We can describe the event that a player is rational, namely, the set of all states in which the player is rational, in terms of her knowledge, as follows. Player  $i$  is *weakly rational* if there is no strategy that she knows yields her a higher payoff than her actual strategy. The player is *strongly rational* if at this state there is no strategy that she knows yields her a payoff no lower than her actual payoff, unless she knows that the strategy yields the same payoff as her actual payoff.

To state this more formally, let  $[\sigma_i \succ \mathbf{s}_i]$  be the event that  $\sigma_i$  yields higher payoffs than the actual payoffs yielded by  $\mathbf{s}_i$ . That is,  $[\sigma_i \succ \mathbf{s}_i] = \{\omega \mid h_i(\sigma_i, \mathbf{s}_{-i}(\omega)) > h_i(\mathbf{s}(\omega))\}$ . Similarly, we define  $[\sigma_i \succeq \mathbf{s}_i] = \{\omega \mid h_i(\sigma_i, \mathbf{s}_{-i}(\omega)) \geq h_i(\mathbf{s}(\omega))\}$ , and  $[\sigma_i \sim \mathbf{s}_i] = \{\omega \mid h_i(\sigma_i, \mathbf{s}_{-i}(\omega)) = h_i(\mathbf{s}(\omega))\}$ .

The following claim follows straightforwardly from Definition 2.

**Claim 1.** *The event that Player  $i$  is weakly rational is*

$$R_i^w = \cap_{\sigma_i \in \Delta(S_i)} \neg K_i[\sigma_i \succ \mathbf{s}_i].$$

*The event that Player  $i$  is strongly rational is*

$$R_i^s = \cap_{\sigma_i \in \Delta(S_i)} (\neg K_i[\sigma_i \succeq \mathbf{s}_i]) \cup (K_i[\sigma_i \sim \mathbf{s}_i]).$$

**3.2. The equivalence theorem.** We define two solutions for games in strategic form that correspond to the two notions of rationality.

**Definition 3.** A non-empty set of strategy profiles  $C \subseteq S$  is a *weak non-probabilistic correlated equilibrium* if for each  $s \in C$  and each player  $i$ ,  $s_i$  is not strongly dominated relative to  $\{t_{-i} \mid (s_i, t_{-i}) \in C\}$ . A non-empty set of strategy profiles  $C \subseteq S$  is a *strong non-probabilistic correlated equilibrium* if for each  $s \in C$  and each player  $i$ ,  $s_i$  is not weakly dominated relative to  $\{t_{-i} \mid (s_i, t_{-i}) \in C\}$ .

Our characterization theorem claims that the family of weak non-probabilistic correlated equilibria is the same as the family of profile sets that are played when there is common knowledge of weak rationality and that the family of strong non-probabilistic correlated equilibria is the same as the family of profile sets that are played when there is common knowledge of strong rationality. To state these results formally, we denote by  $\mathbf{s}(E)$  the set of profiles played at the event  $E$ , that is, the set  $\{\mathbf{s}(\omega) \mid \omega \in E\}$ . The set  $R^w = \cap_{i \in I} R_i^w$  is the event that all players are weakly rational and  $R^s = \cap_{i \in I} R_i^s$  is the event that all players are strongly rational.

**Theorem 1.** *Let  $C$  be a non-empty set of profiles.  $C$  is a weak non-probabilistic correlated equilibrium if and only if  $C = \mathbf{s}(\text{CKR}^w)$  in some knowledge structure for  $G$ .  $C$  is a strong non-probabilistic correlated equilibrium if and only if  $C = \mathbf{s}(\text{CKR}^s)$  in some knowledge structure for  $G$ .*

*Proof.* We prove the second part of the theorem. The proof of the first part is similar. Let  $C$  be a non-empty set of profiles and assume that  $C = \mathbf{s}(\text{CKR}^s)$ . Suppose that for some  $s \in C$  and  $i$ ,  $s_i$  is weakly dominated by  $\sigma_i$  relative to  $\{t_{-i} \mid (s_i, t_{-i}) \in C\}$ . Thus for all  $\omega \in \text{CKR}^s$  such that  $\mathbf{s}_i(\omega) = s_i$ ,  $h_i(\sigma_i, \mathbf{s}_{-i}(\omega)) \geq h_i(\mathbf{s}(\omega))$ , and for one such  $\bar{\omega}$ , strict inequality holds. Since  $\mathbf{s}_i(\bar{\omega}) = s_i$ ,  $\mathbf{s}_i(\omega') = s_i$  for all  $\omega' \in \Pi_i(\bar{\omega})$ . Since  $\bar{\omega} \in \text{CKR}^s$ ,  $\Pi_i(\bar{\omega}) \subseteq \text{CKR}$ , and thus the weak inequality holds for all the states in  $\Pi_i(\bar{\omega})$ . This, with the strong inequality for  $\bar{\omega}$ , imply that Player  $i$  is not strongly rational at  $\bar{\omega}$ . However,  $\bar{\omega} \in \text{CKR}^s \subseteq R_i^s$ , which is a contradiction.

Conversely, suppose that  $C$  is a strong non-probabilistic correlated equilibrium. We construct a knowledge space for which  $\mathbf{s}(\text{CKR}^s) = C$ . The space  $\Omega$  consists of states  $\omega_s$ , one for each  $s \in C$ . The partitions  $\Pi_i$  are defined by,  $\Pi_i(\omega_s) = \{\omega_{\bar{s}} \mid \bar{s} \in C, \bar{s}_i = s_i\}$ . Finally,  $\mathbf{s}(\omega_s) = s$ . Obviously, by the definition of  $C$ , each player  $i$  is strongly rational in each state, and therefore,  $\text{CKR}^s = \Omega$ , and  $\mathbf{s}(\text{CKR}^s) = C$ .  $\square$

**3.3. Domination by a pure strategy.** We have up to this point dealt with a model in which at each state players play a particular pure strategy. When we considered definitions of rationality however we considered the payoff that would result if a player deviated to some mixed strategy. One might well object that if players can only play pure strategies, comparisons should be made only to the result of playing other pure strategies and not mixed ones. There are two possible responses to such an objection. One could develop the theory purely in terms of domination by pure strategies. Or one could provide some justification for considering domination by mixed strategies, even in a model in which players play, in each state, a pure strategy.

The first response would involve redoing the theory developed above, replacing everywhere  $\sigma_i \in \Delta(S_i)$  with  $t_i \in S_i$ . We would thus obtain a theory for domination by *pure* strategies. The notions of weak and strong rationality obtained in this way are weaker than the corresponding notions defined above using domination by mixed strategies. Any weak or strong non-probabilistic correlated equilibrium defined using domination by mixed strategies is a weak or strong non-probabilistic correlated equilibrium defined using domination by pure strategies, but there may be others as well (see Example 4 below which demonstrates that this may happen even for games in extensive form and perfect information). For Bayesian rationality and probabilistic correlated equilibrium, it makes no difference whether players' deviations are pure or mixed, because of the linearity of expectation.

Although the theory for pure strategy domination has weaker predictions it has the advantage that it can be applied to games in which only ordinal preferences of outcomes are given. Assuming mixed strategy domination, and, a fortiori, also assuming Bayesian rationality and correlated equilibrium, requires that payoffs are in terms of a von Neumann Morgenstern utility function.

The second response would justify the theory developed above by showing that nothing is gained if players are allowed to play mixed strategies rather than pure strategies. We sketch this explanation in the next section.

**3.4. Extended structures.** In knowledge structures as defined here, and in knowledge and belief structures used in Aumann (1987), the strategy profile in each state is pure. We call a structure in which a profile of *mixed* strategies is specified in each state, an *extended structure*. For the probabilistic case in Aumann (1987), the theory remains the same for extended structures because of the linearity of expectation. In our case too, despite the absence of linearity, extended models do not add to the theory.

Consider an extended knowledge structure  $(\Omega, (\Pi_i)_{i \in I}, \sigma)$ , where  $\sigma$  assigns a profile of mixed strategies to each state such that  $\sigma_i$  is fixed on each element of  $\Pi_i$ . Weak and strong rationality in this extended model are defined in the same way they are defined for knowledge structure with pure strategy function  $\mathbf{s}$ . For the purpose of this subsection we use  $R^w$  and  $R^s$  to denote the events that all players are weakly or strongly rational in the extended model.

The *support* of a mixed strategy profile  $\sigma$  consists of all the pure strategy profiles  $\mathbf{s}$ , such that  $\sigma_i(s_i) > 0$  for all  $i$ . The support of a set of mixed strategy profiles is the union of the supports of the mixed strategy profiles in the set.

**Proposition 1.** *Let  $C$  be a non-empty subset of pure strategy profiles. Then  $C$  is a weak (strong) non-probabilistic correlated equilibrium if and only if it is the support of  $\sigma(\text{CKR}^w)$  ( $\sigma(\text{CKR}^s)$ ) in some extended knowledge structure of the game.*

We give only a sketch of the proof. One direction is simple, as a knowledge structure is, in particular, an extended one. For the converse, we construct for a given extended knowledge structure a knowledge structure that describes the realizations of the mixed profiles. Thus, each state of the extended structure is replaced by a set of states that correspond to the possible realizations of pure strategies in the state. We assume that each player knows the realizations of her mixed strategy but does not know the realizations of her opponents. Obviously, the set of strategies played in any event  $E$  of the knowledge structure is the support of the mixed profiles played in the corresponding event in the extended structure that gave rise to  $E$ . Also, if Player  $i$  is not rational in a state  $\omega$  in which she plays  $s_i = \mathbf{s}_i(\omega)$  that is dominated by  $\sigma_i$  in the knowledge structure, she could do better in the extended structure in the state from which  $\omega$  originated by replacing  $s_i$  in her mixed strategy by  $\sigma_i$ .

**3.5. Examples.** In the following examples we examine the two notions of non-probabilistic correlated equilibria in two player, 2 by 2 games. For such games domination by pure and mixed strategies are the same.

**Example 1.** Let  $C$  be a strong non-probabilistic correlated equilibrium of the game in Figure 1. Strategy  $r$  is weakly dominated by  $l$ , and against  $b$ ,  $l$  yields a higher



payoff to Player 2 than  $r$ . Thus,  $r$  is weakly dominated by  $l$  relative to any subset of Player's 2 strategies that includes  $b$ . We conclude that  $(b, r) \notin C$ .

		Player 2	
		$l$	$r$
Player 1	$t$	2, 0	0, 0
	$b$	0, 2	2, 1

FIGURE 1. Comparing weak and strong non-probabilistic correlated equilibria.

Assume, now, that  $(b, l) \in C$ . As  $(b, r) \notin C$ , the only strategy  $s$  of Player 2 for which  $(b, s) \in C$  is  $s = l$ , and  $(t, l)$  yields a higher payoff to 1 than  $(b, l)$ . Thus,  $b$  is weakly dominated by  $t$  relative to  $\{l\}$ . Hence,  $(b, l) \notin C$ .

Thus  $C \subseteq \{(t, l), (t, r)\}$ . It is impossible that  $C = \{(t, r)\}$ , because then  $t$  is dominated by  $b$  relative to  $\{r\}$ . However,  $C = \{(t, l), (t, r)\}$  is a strong non-probabilistic correlated equilibrium as  $t$  is not weakly dominated by  $b$  relative to  $\{r, l\}$ , and also  $r$  and  $l$  are not dominated relative to  $\{t\}$ . It is easy to see that  $C = \{(t, l)\}$  is also a strong non-probabilistic correlated equilibrium.

Obviously, each of the two strong non-probabilistic correlated equilibria is also a weak non-probabilistic correlated equilibrium. However, the set of all four strategy profiles is also a weak non-probabilistic correlated equilibrium, as there are no strongly dominated strategy in this game.

Note, that  $(t, l)$  is the only strategy profile that survives iterated elimination of weakly dominated strategy. Thus, strong non-probabilistic correlated equilibria may contain profiles that do not survive iterated elimination of weakly dominated strategy.

In this game both non-probabilistic correlated equilibria are products of subsets of each player's strategies. The following example shows that this is not always the case.

**Example 2.** Consider the game of Figure 2. This game shows that strong non-probabilistic correlated equilibria may not have a product structure and that this is true even for the largest non-probabilistic correlated equilibrium.

		Player 2	
		$l$	$r$
Player 1	$t$	1, 1	1, 1
	$b$	1, 1	0, 0

FIGURE 2. A game with a non-product strong non-probabilistic correlated equilibrium.

The profile  $(b, r)$  cannot be a member of any strong non-probabilistic correlated equilibrium. The argument can be taken verbatim from the previous example. It is easy to see that any non-empty subset of profiles that does not contain  $(b, r)$  is a strong non-probabilistic correlated equilibrium.

Each one of the strong non-probabilistic correlated equilibria is a fortiori a weak non-probabilistic correlated equilibrium. However, the set of all four profiles is also a weak non-probabilistic correlated equilibrium, as there are no strongly dominated strategies in this game. This is the only weak non-probabilistic correlated equilibrium that contains  $(b, r)$ . In any other set of profiles  $C$ , that contains  $(b, r)$ , either  $b$  or  $r$  will be strongly dominated relative to the other player's strategies in  $C$ .

There are three sets of profiles that arise as a result of eliminating weakly dominated strategies:  $\{(t, l), (t, r)\}$ ,  $\{(t, l), (b, l)\}$ , and  $\{(t, l)\}$ . Each of these sets is a strong non-probabilistic correlated equilibrium, however the family of strong non-probabilistic correlated equilibria has four additional members.

As this example demonstrates, weak and strong non-probabilistic correlated equilibria are not necessarily product of sets of players' strategies. But the largest weak non-probabilistic correlated equilibrium in this example is a product, which as we show later holds true for all games. In contrast, the largest strong non-probabilistic correlated equilibrium in this game is not a product. Nevertheless, it contains all the product sets obtained by iterated elimination of weakly dominated strategies. This also holds true for all games, as we show later.

#### 4. PROPERTIES OF NON-PROBABILISTIC CORRELATED EQUILIBRIA

**4.1. The largest non-probabilistic correlated equilibrium.** We show that the families of weak and strong non-probabilistic correlated equilibrium have each a largest element with respect to the partial order of set inclusion. This is done by constructing these largest sets, which also proves the existence of weak and strong non-probabilistic correlated equilibria. The largest non-probabilistic correlated equilibrium is reached by iterated elimination of profiles which we call flaws, and is independent of the order of elimination.

**Definition 4.** A profile  $s$  in a set  $T \subseteq S$  is said to be a *strong flaw* in  $T$  if for some  $i$ ,  $s_i$  is strongly dominated relative to  $\{t_{-i} \mid (s_i, t_{-i}) \in T\}$ . The profile  $s$  is said to be a *weak flaw* in  $T$ , if for some  $i$  and  $\sigma_i$ ,  $s_i$  is weakly dominated by  $\sigma_i$  relative to  $\{t_{-i} \mid (s_i, t_{-i}) \in T\}$ , and  $h_i(\sigma_i, s_{-i}) > h_i(s)$ .

The following two lemmata are used in showing that iterated elimination of flaws results in the largest non-probabilistic correlated equilibria.

**Lemma 1.** *A non-empty set  $T \subseteq S$  is a weak (strong) non-probabilistic correlated equilibrium if and only if there are no strong (weak) flaws in  $T$ .*

*Proof.* This is straightforward for weak non-probabilistic correlated equilibrium and strong flaws. If  $s$  is a weak flaw in  $T$ , then obviously  $T$  is not a strong non-probabilistic correlated equilibrium. If  $T$  is not a weak non-probabilistic correlated equilibrium, then for some  $s \in T$ ,  $i$  and  $\sigma_i$ ,  $s_i$  is weakly dominated by  $\sigma_i$  relative to  $\{t_{-i} \mid (s_i, t_{-i}) \in T\}$ . Thus, for some  $t_{-i}$  in this set,  $h_i(\sigma_i, t_{-i}) > h_i(s)$ . The profile  $(s_i, t_{-i})$  is a weak flaw in  $T$ .  $\square$

**Lemma 2.** *If  $s \in T \subseteq T' \subseteq S$  is a weak (strong) flaw in  $T'$  it is also a weak (strong) flaw in  $T$ .*

*Proof.* Indeed, if  $s$  is a strong flaw in  $T'$ , then for some  $i$ ,  $s_i$  is strongly dominated relative to  $\{t_{-i} \mid (s_i, t_{-i}) \in T'\}$ , and therefore it is also strongly dominated relative to  $\{t_{-i} \mid (s_i, t_{-i}) \in T\}$  which is a subset of the former set. Thus  $s$  is a strong flaw in  $T$ . If  $s_i$  is weak flaw in  $T'$  then for some  $i$  and  $\sigma_i$ ,  $s_i$  is weakly dominated by  $\sigma_i$  relative to  $\{t_{-i} \mid (s_i, t_{-i}) \in T'\}$ , and  $h_i(\sigma_i, s_{-i}) > h_i(s)$ . Since  $\{t_{-i} \mid (s_i, t_{-i}) \in T\}$  is a subset of  $\{t_{-i} \mid (s_i, t_{-i}) \in T'\}$ , and  $s$ , that contributes the strict inequality, is in  $T$ , it follows that  $s$  is a weak flaw in  $T$ .  $\square$

**Theorem 2.** *Consider a sequence of  $m > 0$  subsets of profiles,  $S = S^0 \supset S^1 \supset \dots \supset S^m$ , where for  $0 < k \leq m$ ,  $S^k$  is obtained by eliminating from  $S^{k-1}$  some profiles which are weak (strong) flaws in  $S^{k-1}$ , and where  $S^m$  has no weak (strong) flaws in it. Then,  $S^m$  is the largest strong (weak) non-probabilistic correlated equilibrium.*

*Proof.* We show by induction on  $k$  that if  $C$  is a strong (weak) non-probabilistic correlated equilibrium, then  $C \subseteq S^k$  for  $k = 0, \dots, m$ . This obviously holds for  $k = 0$ . Suppose  $C \subseteq S^k$ . If  $s \in C$ , then it cannot be a strong (weak) flaw in  $S^k$  because in that case it would be, by Lemma 2, a strong (weak) flaw in  $C$ , and by Lemma 1 there are no flaws in  $C$ . Therefore,  $C \subseteq S^{k+1}$ . Since  $S^m$  has no strong (weak) flaw it is a weak (strong) non-probabilistic correlated equilibrium, by Lemma 1. As  $S^m$  contains all weak (strong) non-probabilistic correlated equilibria, it is the largest.  $\square$

The family of correlated equilibria is a convex set. It follows that the family of supports of correlated equilibria is closed under unions. This property holds also for non-probabilistic correlated equilibrium and in particular serves as another proof of the existence of a largest non-probabilistic correlated equilibrium.

**Proposition 2.** *Each of the families of weak and strong non-probabilistic correlated equilibria is closed under unions.*

*Proof.* We prove the contrapositive. Let  $C_1$  and  $C_2$  be two subsets of  $S$ . Suppose that  $C_1 \cup C_2$  is not a weak non-probabilistic correlated equilibrium. Thus there is some  $s \in C_1 \cup C_2$ , some  $i$ , and some  $\sigma_i \in \Delta(S_i)$  such that  $s_i$  is strongly dominated by  $\sigma_i$  relative to  $\{t_{-i} \mid (s_i, t_{-i}) \in C_1 \cup C_2\}$ . Without loss of generality suppose  $s \in C_1$ . Then  $s_i$  is strongly dominated by  $\sigma_i$  relative to  $\{t_{-i} \mid (s_i, t_{-i}) \in C_1\}$  and so  $C_1$  is not a weak non-probabilistic correlated equilibrium.

Suppose that  $C_1 \cup C_2$  is not a strong non-probabilistic correlated equilibrium. Thus there is some  $s \in C_1 \cup C_2$ , some  $i$ , and some  $\sigma_i \in \Delta(S_i)$  such that  $s_i$  is weakly dominated by  $\sigma_i$  relative to  $\{t_{-i} \mid (s_i, t_{-i}) \in C_1 \cup C_2\}$ . Let  $\bar{t}_i$  be such that  $h_i(\sigma_i, \bar{t}_i) > h_i(s_i, \bar{t}_i)$ . Without loss of generality suppose  $(s_i, \bar{t}_i) \in C_1$ . Then  $s_i$  is weakly dominated by  $\sigma_i$  relative to  $\{t_{-i} \mid (s_i, t_{-i}) \in C_1\}$  and so  $C_1$  is not a strong non-probabilistic correlated equilibrium.  $\square$

**4.2. Iterated elimination of dominated strategies.** By Theorem 2, any iterated elimination of flaws results in the maximal non-probabilistic correlated equilibrium, independently of the order of elimination. For the case of elimination of strong flaws we can choose an order that results in sequence of sets that are products. That is, for  $k = 0, \dots, m$ ,  $S^k = \times_{i \in I} S_i^k$ . In step  $k + 1$  we eliminate for each  $i$  all strong flaws  $s$  such that  $s_i$  is strongly dominated relative to  $S_{-i}^k$ . Note, that in this case, all profiles  $(s_i, s_{-i})$ , where  $s_{-i} \in S_{-i}^k$  are strong flaws. Thus,  $S^{k+1}$  is

the product resulted from elimination of strongly dominated strategies in  $S^k$ . We conclude:

**Theorem 3.** *The largest weak non-probabilistic correlated equilibrium is the set of profiles that survive iterated elimination of strongly dominated strategies.*

While iterated elimination of strongly dominated strategies results in the same set of profiles, independently of the order of elimination, the result of iterated elimination of weakly dominated strategies may depend on the order of elimination (see, for example, Myerson, 1991, 57–61). Thus, we cannot expect a claim similar to Theorem 3 for strong non-probabilistic correlated equilibrium. Observe also that in the iterated elimination of weak flaws, as opposed to the case of strong flaws, whole strategies are not necessarily eliminated. Thus, the end result, the largest strong non-probabilistic correlated equilibrium, can be a set which is not a product, as demonstrated by Example 2. However, we can state the following relationship of iterated elimination of weakly dominated strategies and strong non-probabilistic correlated equilibrium.

**Theorem 4.** *Any set of profiles that survive iterated elimination of weakly dominated strategies is a strong non-probabilistic correlated equilibrium.*

We shall prove this theorem using the following lemma.

**Lemma 3.** *If a strategy  $\sigma_i \in \Delta(S_i)$  is weakly dominated, then it is weakly dominated by a strategy  $\sigma'_i \in \Delta(S_i^u)$ , where  $S_i^u$  is the set of strategies in  $S_i$  which are not weakly dominated.*

*Proof.* Let  $S_i^0$  be a minimal set of strategies in  $S_i$  such that each weakly dominated strategy in  $\Delta(S_i)$  is weakly dominated by a strategy in  $\Delta(S_i^0)$ . Suppose that a weakly dominated strategy  $t_i$  is in  $S_i^0$ . Then,  $t_i$  is weakly dominated by some strategy  $\alpha t_i + (1 - \alpha)\sigma_i$ , where  $\sigma_i \in \Delta(S_i^0 \setminus \{t_i\})$ . This implies that  $t_i$  is weakly dominated by  $\sigma_i$ . Now, for any strategy which is weakly dominated there is strategy  $\alpha t_i + (1 - \alpha)\sigma'_i$ , where  $\sigma'_i \in \Delta(S_i^0 \setminus \{t_i\})$ , that weakly dominates it. But this strategy is weakly dominated by  $\alpha \sigma_i + (1 - \alpha)\sigma'_i$ , which is a contradiction to the minimality of  $S_i^0$ .  $\square$

*Proof of Theorem 4.* Let  $S' = \times_{i \in I} S'_i$  be a subset of  $S$  obtained by one round of elimination of weakly dominated strategies in  $S$ , and let  $G'$  be the game whose set of strategy profiles is  $S'$ . We show that if  $C \subseteq S'$  is a strong non-probabilistic correlated equilibrium in the game  $G'$  it is also a strong non-probabilistic correlated equilibrium in  $G$ . If this is not the case, then for some  $s \in C$  and player  $i$ ,  $s_i$  is weakly dominated by some  $\sigma_i \in \Delta(S_i)$  relative to  $\{t_{-i} \mid (s_i, t_{-i}) \in C\}$ . Now, any  $\sigma_i \in \Delta(S_i)$  either belongs to  $\Delta(S'_i)$ , or else it is weakly dominated, and hence, by Lemma 3, it is weakly dominated by some  $\sigma'_i \in \Delta(S_i^u) \subseteq \Delta(S'_i)$ . In either case,  $s_i$  is weakly dominated relative to the said set by a strategy in  $\Delta(S'_i)$ , contrary to our assumption that  $C$  is a non-probabilistic correlated equilibrium in  $G'$ .

Obviously, if a set of profiles  $C$  is obtained by iterated elimination of weakly dominated strategies, then  $C$  is a strong non-probabilistic correlated equilibrium in the game whose set of strategy profiles is  $C$ . Thus, applying the previous claim iteratively we reach the desired conclusion.  $\square$

**4.3. The centipede game.** We consider a centipede game with  $n$  stages. Player 1 chooses in the odd stages between going down and terminating the game, or across, and Player 2, in the even stages. We denote the player who chooses at stage  $k$  as  $i(k)$ , the other player as  $j(k)$ , the outcome or terminal vertex after  $i(k)$  chooses down as  $z_k$ , and the terminal vertex reached if no player chooses down as  $z_{n+1}$ . The payoffs at  $z_k$  are  $k + 2$  to  $i(k)$  and  $k$  to  $j(k)$ , that is, the payoffs are  $(k + 2, k)$  if  $k$  is odd and  $(k, k + 2)$  if  $k$  is even. The game tree is given in Figure 3.

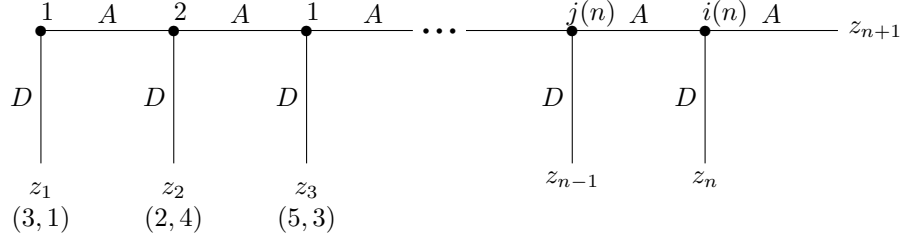


FIGURE 3. The  $n$  Stage Centipede Game in extensive form.

If  $n$  is even then each player has  $n/2 + 1$  classes of Kuhn-equivalent strategies that we label  $1, 3, 5, \dots, n + 1$  for Player 1 and  $2, 4, 6, \dots, n + 2$  for Player 2. If  $n$  is odd then Player 1 has  $(n + 1)/2 + 1$  classes of Kuhn-equivalent strategies and Player 2 has  $(n + 1)/2$  classes of Kuhn-equivalent strategies that we label  $1, 3, 5, \dots, n + 2$  for Player 1 and  $2, 4, 6, \dots, n + 1$  for Player 2. The class  $k$ , for  $k = 1, \dots, n$ , contains all strategies in which a player  $i(k)$  chooses down for the first time at stage  $k$ . The classes  $n + 1$  and  $n + 2$  consist of the strategy of always playing across. We refer to these classes as strategies. The strategy pair  $(s_1, s_2)$  induces the outcome  $z_k$  where  $k = \min\{s_1, s_2\}$ .

The strategic form of the game for the case of is given in Figure 4.

		Player 2					
		2	4	6	...	$n$	$n + 2$
Player 1	1	$\begin{smallmatrix} z_1 \\ (3, 1) \end{smallmatrix}$	$\begin{smallmatrix} z_1 \\ (3, 1) \end{smallmatrix}$	$\begin{smallmatrix} z_1 \\ (3, 1) \end{smallmatrix}$		$\begin{smallmatrix} z_1 \\ (3, 1) \end{smallmatrix}$	$\begin{smallmatrix} z_1 \\ (3, 1) \end{smallmatrix}$
	3	$\begin{smallmatrix} z_2 \\ (2, 4) \end{smallmatrix}$	$\begin{smallmatrix} z_3 \\ (5, 3) \end{smallmatrix}$	$\begin{smallmatrix} z_3 \\ (5, 3) \end{smallmatrix}$		$\begin{smallmatrix} z_3 \\ (5, 3) \end{smallmatrix}$	$\begin{smallmatrix} z_3 \\ (5, 3) \end{smallmatrix}$
	5	$\begin{smallmatrix} z_2 \\ (2, 4) \end{smallmatrix}$	$\begin{smallmatrix} z_4 \\ (4, 6) \end{smallmatrix}$	$\begin{smallmatrix} z_5 \\ (7, 5) \end{smallmatrix}$		$\begin{smallmatrix} z_5 \\ (7, 5) \end{smallmatrix}$	$\begin{smallmatrix} z_5 \\ (7, 5) \end{smallmatrix}$
	$\vdots$						
	$n - 1$	$\begin{smallmatrix} z_2 \\ (2, 4) \end{smallmatrix}$	$\begin{smallmatrix} z_4 \\ (4, 6) \end{smallmatrix}$	$\begin{smallmatrix} z_6 \\ (6, 8) \end{smallmatrix}$		$z_{n-1}$	$z_{n-1}$
	$n + 1$	$\begin{smallmatrix} z_2 \\ (2, 4) \end{smallmatrix}$	$\begin{smallmatrix} z_4 \\ (4, 6) \end{smallmatrix}$	$\begin{smallmatrix} z_6 \\ (6, 8) \end{smallmatrix}$		$z_n$	$z_{n+1}$

FIGURE 4. The  $n$  Stage Centipede Game in strategic form for even  $n$ .

**Proposition 3.** *In the centipede game, a set  $C$  of profiles is a strong non-probabilistic correlated equilibrium if and only if  $(1, 2) \in C$  and  $C$  contains only profiles of the form  $(1, k)$ .*

*Proof.* Let  $C$  be a strong non-probabilistic correlated equilibrium. There is a  $\bar{k}$  such that all profiles in  $C$  induce an outcome  $z_k$  with  $k \leq \bar{k}$ , namely  $\bar{k} = n + 1$ . Let  $\underline{k}$  be the smallest such  $\bar{k}$ . We claim that  $\underline{k} = 1$ . Suppose that  $\underline{k} > 1$ . There is some profile  $(s_1, s_2)$  in  $C$  that induces the outcome  $z_{\underline{k}}$ . This profile must have  $s_{i(\underline{k})} = \underline{k}$  and  $s_{j(\underline{k})} > \underline{k}$ . (Otherwise the outcome would not be  $z_{\underline{k}}$ .) Also there are no profiles  $(t_1, t_2)$  in  $C$  with  $t_{j(\underline{k})} = s_{j(\underline{k})}$  and  $t_{i(\underline{k})} > \underline{k}$ , since then there would be a profile inducing an outcome  $z_k$  with  $k > \underline{k}$ . But now strategy  $s_{j(\underline{k})}$  is weakly dominated by  $\underline{k} - 1$  relative to  $\{t_{i(\underline{k})} \mid (t_1, t_2) \in C, t_{j(\underline{k})} = s_{j(\underline{k})}\}$  contradicting the assumption that  $C$  is a strong non-probabilistic correlated equilibrium.

Thus all members of  $C$  must induce the outcome  $z_1$  and so Player 1's strategy must be 1 in all profiles in  $C$ . When Player 1's strategy is 1, Player 2's payoff is fixed, and thus no strategy of hers is dominated relative to strategy 1. If  $(1, 2) \in C$  then, since in this profile Player 1 strictly prefers 1 to any other strategy his strategy is not dominated relative to any subset of Player 2's strategies. However if  $(1, 2)$  is not in  $C$  then 1 is strictly dominated by 3 for Player 1 and so  $C$  is not a non-probabilistic correlated equilibrium.  $\square$

There are no strongly dominated strategies in the centipede game, and thus the set of all profiles is a weak non-probabilistic correlated equilibrium. Thus, common knowledge of weak rationality does not imply much for this game. Strong non-probabilistic correlated equilibria are more restricted and predict immediate termination of the game by Player 1. An iterated elimination of weakly dominated strategies is even more restrictive than strong non-probabilistic correlated equilibrium: it results in a single profile  $(1, 2)$  which, unlike strong non-probabilistic correlated equilibrium, determines not only Player 1's strategy, but also Player 2's strategy.

**4.4. Probabilistic and non-probabilistic correlated equilibria.** A (*probabilistic*) *correlated equilibrium* of the game  $G$  is a probability distribution  $p \in \Delta(S)$  such that for all  $i$ , and all  $s_i$  and  $t_i$  in  $S_i$ ,

$$\sum_{s_{-i} \in S_{-i}} p(s_i, s_{-i}) h_i(s_i, s_{-i}) \geq \sum_{s_{-i} \in S_{-i}} p(s_i, s_{-i}) h_i(t_i, s_{-i}).$$

Observe that by the linearity of expectation, these inequalities hold also if we change  $t_i$  to a mixed strategy  $\sigma_i \in \Delta(S_i)$ .

**Proposition 4.** *The support of a correlated equilibrium is a strong non-probabilistic correlated equilibrium.*

*Proof.* Let  $p \in \Delta(S)$  be a correlated equilibrium, the support of which is  $C$ . Suppose that for  $s \in C$ ,  $s_i$  is weakly dominated by  $\sigma_i$  relative to  $R = \{t_{-i} \mid (s_i, t_{-i}) \in C\}$ . Then, for all  $t_{-i} \in R$ ,  $h_i(s_i, t_{-i}) \geq h_i(\sigma_i, t_{-i})$  with at least one strict inequality. By the definition of  $C$ ,  $p(s_i, t_{-i}) > 0$  if and only if  $t_{-i} \in R$ . Therefore,

$$\begin{aligned}
\sum_{t_{-i} \in S_{-i}} p(s_i, t_{-i}) h_i(s_i, t_{-i}) &= \sum_{t_{-i} \in R} p(s_i, t_{-i}) h_i(s_i, t_{-i}) \\
&> \sum_{t_{-i} \in R} p(s_i, t_{-i}) h_i(\sigma_i, t_{-i}) \\
&= \sum_{t_{-i} \in S_{-i}} p(s_i, t_{-i}) h_i(\sigma_i, t_{-i}),
\end{aligned}$$

contrary to our assumption that  $p$  is a correlated equilibrium.  $\square$

The converse of Proposition 4 does not hold. That is, a non-probabilistic correlated equilibrium is not necessarily the support of a correlated equilibrium. Moreover, a non-probabilistic correlated equilibrium can be disjoint from the largest support of a correlated equilibrium, as we show in the next example.

**Example 3.** The game in Figure 5 is a two person zero sum game. Therefore, for a correlated equilibrium, the conditional probability in each column, when the column has a positive probability, is an optimal strategy of Player 1 (see Forges, 1990). The value of the game is 1, and the only optimal strategy of player 1 is  $B$ . Thus, the support of any correlated equilibrium is in  $\{(B, H), (B, T)\}$ . Indeed, any probability distribution on this set is a correlated equilibrium. However, it is easy to see that the set of the four top profiles is a non-probabilistic correlated equilibrium.

		Player 2	
		$H$	$T$
Player 1	$H$	4, -4	-4, 4
	$T$	-4, 4	4, -4
	$B$	1, -1	1, -1

FIGURE 5. Non-probabilistic and probabilistic correlated equilibria.

## 5. GAMES WITH PERFECT INFORMATION

In this section the game  $G$  is the strategic form of a game in extensive form with perfect information. For such games the notions of rationality in Definition 2 can be strengthened by requiring rationality in subgames  $G^v$  that start at vertices  $v$ . It might seem that in games with perfect information it makes no difference whether one defines the concepts using domination by pure strategies or by mixed strategies. This is, however, not true, as the next example shows.

**Example 4.** Consider the extensive form game with perfect information given in Figure 6 with the associated strategic form given in Figure 7.

Consider the set  $C = \{(B, HH), (B, TT)\}$ . Relative to  $\{B\}$  all of Player 2's strategies are equivalent so none are dominated relative to  $\{B\}$ . And for Player 1,  $B$

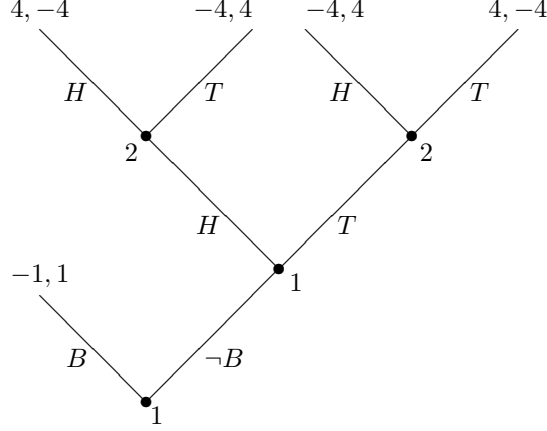


FIGURE 6. Domination by pure or mixed strategies in a game with perfect information. Extensive Form.

		Player 2			
		$HH$	$HT$	$TH$	$TT$
Player 1	$H$	4, -4	4, -4	-4, 4	-4, 4
	$T$	-4, 4	4, -4	-4, 4	4, -4
	$B$	-1, 1	-1, 1	-1, 1	-1, 1

FIGURE 7. Domination by pure or mixed strategies in a game with perfect information. Strategic Form.

is not weakly dominated by either  $H$  or  $T$  relative to  $\{HH, TT\}$ ; so if dominance is defined as domination by a pure strategy  $C$  is a strong non-probabilistic correlated equilibrium. However,  $B$  is (even strongly) dominated relative to  $\{HH, TT\}$  by the mixed strategy that puts weight a half on each of  $H$  and  $T$ ; so if dominance is defined as domination by a mixed strategy  $C$  is not a strong non-probabilistic correlated equilibrium. It is worth pointing out that, even with dominance defined as domination by a mixed strategy there is a strong non-probabilistic correlated equilibrium in which Player 1 plays  $B$ , namely  $\{(B, TH)\}$ . It is straightforward to see this is a strong non-probabilistic correlated equilibrium. In fact,  $(B, TH)$  is a pure strategy Nash equilibrium.

Aumann (1995) and Aumann (1998) studied two notions of rationality, substantive and material rationality, correspondingly. Rationality is defined in these papers by pure strategy domination. The results reported there cannot be strengthened by



assuming mixed strategy domination. Following these papers we assume in this section that weak and strong rationality, as well as weak and strong non-probabilistic correlated equilibria are defined in terms of pure strategy domination.

**5.1. Substantive rationality.** We recall that a knowledge structure for  $G$  is a knowledge structure for the games  $G^v$  for all vertices  $v$ .

**Definition 5.** Player  $i$  is *substantively weakly (strongly) rational* in a knowledge structure for  $G$  if for each vertex  $v$ ,  $i$  is weakly (strongly) rational when it is viewed as a knowledge structure for the game  $G^v$ .

We denote by  $[t_i^v \succ s_i^v]$  the event that  $i$ 's strategy in  $G^v$ ,  $t_i^v$ , yields higher payoffs than yielded by  $s_i^v$  in this game. That is,  $[t_i^v \succ s_i^v] = \{\omega \mid h_i^v(t_i^v, s_{-i}^v(\omega)) > h_i^v(s_i^v(\omega))\}$ . The events  $[t_i^v \succeq s_i^v]$  and  $[t_i^v \sim s_i^v]$  are similarly defined. Then, the event that  $i$  is substantively weakly rational is

$$R_i^{\text{sw}} = \bigcap_{v \in V} \bigcap_{t_i^v \in S_i^v} \neg K_i[t_i^v \succ s_i^v],$$

and the event that  $i$  is substantively strongly rational is

$$R_i^{\text{ss}} = \bigcap_{v \in V} \bigcap_{t_i^v \in S_i^v} (\neg K_i[t_i^v \succeq s_i^v]) \cup (K_i[t_i^v \sim s_i^v]).^3$$

To describe the set of strategy profiles played when there is common knowledge of substantive rationality, we define the following solution.

**Definition 6.** A set of strategy profiles  $C \subseteq S$  is a *non-probabilistic substantive weak (strong) correlated equilibrium*, if for each  $v$ ,  $C^v$  is a non-probabilistic correlated equilibrium for the game  $G^v$ .

Denote by  $R^{\text{sw}}$  and  $R^{\text{ss}}$  the events that all players are substantively weakly and strongly rational, correspondingly. The following theorem follows straightforwardly from the definitions and from Theorem 1, and the proof is omitted.

**Theorem 5.** Let  $C$  be a non-empty set of profiles. Then,  $C$  is a non-probabilistic substantive weak (strong) correlated equilibrium if and only if  $C = \mathbf{s}(\text{CK}R^{\text{sw}})$  ( $C = \mathbf{s}(\text{CK}R^{\text{ss}})$ ) in some knowledge structure for  $G$ .

Of interest are games in extensive form in general position. For such games it is straightforward to see:

**Claim 2.** A game in extensive form with perfect information in a general position has a unique non-probabilistic substantive weak (and a fortiori, a unique strong) correlated equilibrium, which is the singleton that consists of the unique backward induction profile.

Aumann (1995) has shown that common knowledge of substantive weak rationality implies the backward induction outcome. This follows from the general characterization in Theorem 5 and the observation in Claim 2.

<sup>3</sup>The intersections in  $R_i^{\text{sw}}$  and  $R_i^{\text{ss}}$  are taken over all vertices  $v \in V$ . However, it is easy to see that taking intersections over  $i$ 's vertices only, results in the same events. This is how the event  $R_i^{\text{sw}}$  is defined (under the name rationality) in Aumann (1995).

**5.2. Material rationality.** Substantive rationality has a weird feature. It is tested by comparing payoffs in each game  $G^v$ , even if the payoffs are not incurred in the game  $G$  because  $v$  is not reached. Thus, to be substantively rational in a state  $\omega$ , Player  $i$ 's strategy in  $\Pi_i(\omega)$  should not be dominated relative to  $i$ 's opponents' profiles in *all* the states in  $\Pi_i(\omega)$ , even those at which  $v$  is not reached. It is even possible, for example, that  $i$  *knows* that  $v$  is not reached and cannot be reached by her, and still be considered not rational in virtue of her strategy in  $G^v$  that she knows is payoff irrelevant.<sup>4</sup>

This feature of substantive rationality is fixed in material rationality by requiring that a player's strategy in  $G^v$  is not dominated relative to the player's opponents profiles  $i$ 's only in states in which  $v$  is reached, if this set is not empty. For the formal definition, let  $\Omega^v$  be the event that vertex  $v$  is reached. That is,  $\Omega^v$  is the set of all states  $\omega$  such that  $v$  is on the path generated by  $\mathbf{s}(\omega)$ .

**Definition 7.** Player  $i$  is *materially weakly (strongly) rational* at  $\omega$ , if for each  $v \in V$ , either  $\Pi_i(\omega) \subseteq \neg\Omega^v$  or else,  $\mathbf{s}_i^v$  is not strongly (weakly) dominated relative to  $\Pi_i(\omega) \cap \Omega^v$ .

Thus, Player  $i$  is materially weakly rational if for each  $v$  either  $i$  knows that  $v$  is not reached, or else, there is no strategy in  $G^v$  that she knows yields her a higher payoff when  $v$  is reached. Material strong rational is similarly defined. Formally,

**Claim 3.** *The event that  $i$  is materially weakly rational is*

$$R_i^{\text{mw}} = \bigcap_{v \in V} K_i(\neg\Omega^v) \cup \left( \bigcap_{t_i^v \in S_i^v} (\neg K_i(\neg\Omega^v \cup [t_i^v \succ \mathbf{s}_i^v])) \right).$$

*The event that  $i$  is materially strongly rational is*

$$R_i^{\text{ms}} = \bigcap_{v \in V} (K_i(\neg\Omega^v) \cup \left( \bigcap_{t_i^v \in S_i^v} \neg K_i(\neg\Omega^v \cup [t_i^v \succeq \mathbf{s}_i^v]) \cup K_i(\neg\Omega^v \cup [t_i^v \sim \mathbf{s}_i^v]) \right)).$$

Note that for the root vertex  $r$ ,  $\Omega^r = \Omega$ . Thus, the event that  $i$  is weakly (strongly) rational is a superset of the event that she is materially weakly (strongly) rational. That is, material weak (strong) rationality implies weak (strong) rationality. However, for the strong rationality the inverse implication also holds, and therefore:

**Proposition 5.** *A player is materially strongly rational if and only if the player is strongly rational.*

*Proof.* If  $i$  is materially strongly rational, then she is strongly rational. Suppose, now, that  $i$  is not materially strongly rational at  $\omega$ . Then for some  $v$ , there exists  $t_i^v \in S_i^v$  such that  $h_i^v(t_i^v, \mathbf{s}_{-i}^v(\omega')) \geq h_i^v(\mathbf{s}_i^v(\omega'))$  for all states  $\omega' \in \Pi_i(\omega) \cap \Omega^v$ , and at least for one such  $\omega'$  the inequality is strict. The strategy  $t_i$  obtained by changing  $\mathbf{s}_i(\omega)$  to be  $t_i^v$  in the subtree that starts at  $v$ , yield's  $i$  payoffs that are as high as  $\mathbf{s}$  in all states in  $\Pi_i(\omega)$ , and at least in one state in  $\Pi_i(\omega)$ , a higher payoff. This means that  $i$  is not strongly rational at  $\omega$ .  $\square$

<sup>4</sup>To justify this feature of substantive rationality, one can argue that payoff in  $G^v$  when  $v$  is not reached are taken into account counterfactually. This could justify claims like “had  $v$  been reached, strategy  $\mathbf{s}_i^v$  would have been a rational choice”. However, such a statement cannot be formalized in this model. Substantive rationality states instead “ $i$  is rational because have  $v$  been reached, she would have chosen strategy  $\mathbf{s}_i^v$ ”. For a rigorous formulation of strategies as counterfactual statement see Samet (1996) and Di Tillio et al. (2012).

Thus, for games with perfect information we have three notions of rationality with comparable strength, as follows:

$$\begin{array}{ccccccc} & & \text{material} & & \text{material} & & \\ & & \text{strong} & \Rightarrow & \text{weak} & \Rightarrow & \text{weak} \\ \text{strong} & \Leftrightarrow & \text{strong} & \Rightarrow & \text{weak} & \Rightarrow & \text{weak} \\ \text{rationality} & & \text{rationality} & & \text{rationality} & & \text{rationality} \end{array}$$

In light of Proposition 5 we need to study common knowledge of material weak rationality only. We show that it characterizes the following solution.

**Definition 8.** A non-empty set of strategy profile  $C$  is a *materially weak non-probabilistic correlated equilibrium*, if for each vertex  $v \in V$  that is reached by some profile in  $C$ , the subset of profiles in  $C$  that reach  $v$  is a weak non-probabilistic correlated equilibrium in the game  $G^v$ .

The event that all players are materially weakly rational is  $R^{\text{mw}} = \cap_{i \in I} R_i^{\text{mw}}$ .

**Theorem 6.** A non-empty set of profiles  $C$  is materially weak non-probabilistic correlated equilibrium if and only if for some knowledge structure of the game,  $C = \mathbf{s}(\text{CKR}^{\text{mw}})$ .

*Proof.* Let  $C$  be a non-empty set of profiles such that  $C = \mathbf{s}(\text{CKR}^{\text{mw}})$ . Suppose that  $C$  is not a materially weak non-probabilistic correlated equilibrium. Then, for some vertex  $v \in V$  that is reached by a profile in  $C$ , the subset of profiles in  $C$  that reach  $v$ , denoted  $D$ , is not a weak non-probabilistic correlated equilibrium in the game  $G^v$ . Thus, for some  $i$ ,  $s \in D$ , and  $t_i \in S_i$ ,  $t_i^v$  weakly dominates  $s_i^v$  relative to  $\{t_{-i}^v \mid (s_i, t_{-i}) \in D\}$ , and  $h_i^v(t_i, s_{-i}) > h_i^v(s)$ . There exist a state  $\omega$  such that  $s = \mathbf{s}(\omega)$ . As  $\omega \in \Omega^v$ ,  $\Pi_i(\omega)$  is not a subset of  $\neg\Omega^v$ . Moreover,  $\mathbf{s}_{-i}^v(\Pi_i(\omega) \cap \Omega^v) \subseteq \{t_{-i}^v \mid (s_i, t_{-i}) \in D\}$ . Therefore,  $s_i^v = \mathbf{s}_i^v(\omega)$  is weakly dominated by  $t_i^v$  relative to  $\Pi_i(\omega) \cap \Omega^v$ . Hence, each of the two conditions that make  $i$  materially strongly rational at  $\omega$  fail, contrary to our assumption.

Conversely, suppose that  $C$  is a materially strong non-probabilistic correlated equilibrium. We construct a knowledge space the states of which are the profiles in  $C$ , exactly as in the proof of Theorem 1. It is easy to check that each state of this space all players are materially weakly rational. Thus, the whole space is the event  $\text{CKR}^{\text{mw}}$ . By construction, the set of profiles played in the space is  $C$ .  $\square$

For games with perfect information in general position there are only two notions of rationality, as the first three notions collapse into one:

**Proposition 6.** In a game with perfect information in general position, a player is materially weakly rational if and only if the player is strongly rational.

*Proof.* By Proposition 5, strong rationality implies material weak rationality. Conversely, suppose that Player  $i$  is not strongly rational at  $\omega$  in a knowledge structure of such a game. Then for some  $t_i$ ,  $h_i(t_i, \mathbf{s}_{-i}(\omega')) \geq h_i(\mathbf{s}(\omega'))$  for all  $\omega' \in \Pi_i(\omega)$ , and  $h_i(t_i, \mathbf{s}_{-i}(\bar{\omega})) > h_i(\mathbf{s}(\bar{\omega}))$  for some  $\bar{\omega} \in \Pi_i(\omega)$ . The last inequality implies that at some  $v$  reached by  $\mathbf{s}(\bar{\omega})$ , the strategies  $\mathbf{s}_i(\bar{\omega})$  and  $t_i$  choose a different action at  $v$ . The set  $\Pi_i(\omega) \cap \Omega^v$  is not empty, and for all states  $\hat{\omega}$  in this set,  $h_i(t_i, \mathbf{s}_{-i}(\hat{\omega})) \geq h_i(\mathbf{s}(\hat{\omega}))$ . However,  $h_i(\mathbf{s}(\hat{\omega}))$  and  $h_i(t_i, \mathbf{s}_{-i}(\hat{\omega}))$  are incurred in a different subtrees, and thus strong inequality must hold. This means that  $i$  is not materially weakly rational at  $\omega$ .  $\square$

The following is a corollary of Theorem 1, and Proposition 6.

**Corollary 1.** *In a game with perfect information in general position, the set of profiles played when materially weakly rationality is common knowledge is a strong non-probabilistic correlated equilibrium.*

The centipede game is a game in a general position. Thus, we can conclude from Corollary 1 and Proposition 3, that in this game common knowledge of material weak rationality implies that the first player terminates the game immediately. This is the main result of Aumann (1998). However, the immediate termination of the game by Player 1, which is the outcome of the unique Nash equilibrium, and the outcome that holds with probability 1 in all the correlated equilibria of the game, is described there as the outcome of backward induction. Since strong non-probabilistic correlated equilibrium reflects equilibrium rather than backward induction considerations, we conclude by Corollary 1, that backward induction is not an essential feature of common knowledge of material weak rationality. The fact that it implies the backward induction in the centipede game is the result of the coincidence of the backward induction outcome with the outcome in all the profiles in the strong non-probabilistic correlated equilibria of this game.

The notion of *ex post material rationality* defined in Aumann (1998) differs slightly and insignificantly from material weak rationality. Denote by  $R_i^{\text{epm}}$  the event that  $i$  is ex post materially rational. This event differs from  $R_i^{\text{mw}}$  in that the term  $K_i(\neg\Omega^v)$ , in the definition of the latter event, is replaced by  $\neg\Omega^v$ . Therefore,  $R_i^{\text{mw}} = K_i R_i^{\text{epm}}$ . Since we are interested in common knowledge of rationality, the difference between ex post material rationality and material weak rationality is insignificant: common knowledge of both is one and the same event.<sup>5</sup>

## 6. WHEN PLAYERS ERR

We examine now models in which rationality is defined in terms of players' beliefs, rather than knowledge, which enables them to err, and where common knowledge of rationality is replaced by common belief of rationality.

We first introduce contexts of the game  $G$  in which non-probabilistic belief can be expressed. A *belief structure* for  $G$  is a tuple  $(\Omega, (\Pi_i)_{i \in I}, (b_i)_{i \in I}, \mathbf{s})$ , where  $\Pi_i$  is a partition of  $\Omega$ ,  $b_i$  is a function,  $b_i: \Omega \rightarrow 2^\Omega \setminus \{\emptyset\}$ , which is constant on each element of  $\Pi_i$  and for each  $\omega$ ,  $b_i(\omega) \subseteq \Pi_i(\omega)$ , and  $\mathbf{s}: \Omega \rightarrow S$ . We think of  $b_i(\omega)$  as the set of states that are considered possible by  $i$  at  $\omega$ . By the constancy 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  $b_i(\omega)$ .

For each  $i$  the belief operators  $B_i$  is defined by  $B_i E = \{\omega \mid b_i(\omega) \subseteq E\}$ . The common belief operator is defined similarly to the common knowledge operator as follows. We denote by  $BE$  the event that all believe  $E$ , that is,  $BE = \cap_i B_i E$ , and by  $B^n$ , the  $n$ th power of  $B$ . The common belief operator is defined by  $CBE = \cap_{n \geq 1} B^n E$ . Unlike the event  $CKE$  which is a union of elements of  $\Pi_i$  for each  $i$ ,  $CBE$  is not such a union. However, for each  $i$ ,  $CBE \subseteq B_i CBE$  (see, Monderer and Samet, 1989, Proposition 3).<sup>6</sup>

<sup>5</sup>The event  $R_i^{\text{epm}}$  is defined in Aumann (1998) in terms of ex-post knowledge operators  $K_i^v$  for each  $v$ . However, for each  $E$ ,  $K_i^v E = (\Omega^v \cap K_i(\neg\Omega^v \cup E)) \cup (\neg\Omega^v \cap K_i(\Omega^v \cup E))$ , and thus,  $R_i^{\text{epm}}$  can be defined in terms of the operator  $K_i$  alone. Replacing  $K_i^v$  with the equivalent term, in the definition in Aumann (1998), results in the definition given here. See Samet (2011).

<sup>6</sup>Monderer and Samet (1989) show this for  $p$ -belief operators  $B_i^p$ . However, it is easy to check that the properties required to prove the claim are satisfied by the operators  $B_i$ . Alternatively,

A belief structure gives rise also to knowledge operators  $K_i$ , defined by the partitions  $\Pi_i$ . It is easy to see that for each  $i$  and  $E$ ,  $K_i E \subseteq B_i E$ , and  $B_i E \subseteq K_i B_i E$ . Thus, knowledge implies belief, and belief implies the knowledge of the belief. If for each  $\omega$ ,  $b_i(\omega) = \Pi_i(\omega)$ , then  $B_i = K_i$ .<sup>7</sup> Belief structures are discussed, axiomatically defined, and game theoretically applied in Bonanno and Nehring (1998) and Samet (2013).

The events  $R_i^{\text{wb}}$  and  $R_i^{\text{sb}}$ , that Player  $i$  is *doxastically weakly rational*, or *doxastically strongly rational*, correspondingly, are defined by replacing  $K_i$  with  $B_i$  in the events  $R_i^{\text{w}}$  and  $R_i^{\text{s}}$  that  $i$  is (epistemically) weakly or strongly rational.

$$R_i^{\text{wb}} = \cap_{\sigma_i \in \Delta(S_i)} \neg B_i[\sigma_i \succ \mathbf{s}_i].$$

$$R_i^{\text{sb}} = \cap_{\sigma_i \in \Delta(S_i)} \neg B_i[\sigma_i \succeq \mathbf{s}_i] \cup B_i[\sigma_i \sim \mathbf{s}_i].$$

We now give equivalent descriptions of these types of rationality in terms of the belief functions  $b_i$ . Player  $i$  is doxastically weakly (strongly) rational at  $\omega$  if  $\mathbf{s}_i(\omega)$  is not strongly (weakly) dominated relative to the opponents' strategies in the states of  $b_i(\omega)$ . Note, that as knowledge implies belief,  $R_i^{\text{wb}} \subseteq R_i^{\text{w}}$ , that is, doxastic weak rationality implies epistemic weak rationality. However,  $R_i^{\text{sb}}$  is incomparable to  $R_i^{\text{s}}$ .

The characterization of weak non-probabilistic correlated equilibrium in terms of epistemic weak rationality, in Theorem 1, holds mutatis mutandis for doxastic weak rationality.

**Theorem 7.** *Let  $C$  be a non-empty set of profiles. Then,  $C$  is a weak non-probabilistic correlated equilibrium if and only if  $C = \mathbf{s}(\text{CBR}^{\text{wb}})$  in some belief space for  $G$ .*

*Proof.* For each  $i$ ,  $\text{CBR}^{\text{wb}} \subseteq B_i R^{\text{wb}} \subseteq B_i R_i^{\text{wb}}$ , where the first inclusion follows from the definition of common belief, and the second by the monotonicity of  $B_i$ . If  $\omega \in B_i R_i^{\text{wb}}$ , then  $b_i(\omega) \subseteq R_i^{\text{wb}}$  and thus for  $\omega' \in b_i(\omega)$ ,  $\mathbf{s}_i(\omega')$  is not strongly dominated relative to the profiles of the players other than  $i$  in  $b_i(\omega)$ . As  $\mathbf{s}_i(\omega) = \mathbf{s}_i(\omega')$ , it follows that  $i$  is doxastically weakly rational at  $\omega$ . Thus,  $\text{CBR}^{\text{wb}} \subseteq R_i^{\text{wb}}$ .

Suppose that a non-empty set  $C$  satisfies,  $C = \mathbf{s}(\text{CBR}^{\text{wb}})$  and let  $\hat{\Omega} = \text{CKR}^{\text{wb}}$ . Then, by Monderer and Samet (1989),  $\hat{\Omega} \subseteq B_i(\hat{\Omega})$  for each  $i$ . Thus, for each  $\omega \in \hat{\Omega}$ , and each  $i$ ,  $b_i(\omega) \subseteq \hat{\Omega}$ . Consider the *reduced* belief space  $(\hat{\Omega}, (\hat{\Pi}_i)_{i \in I}, (\hat{b}_i)_{i \in I}, \hat{\mathbf{s}})$ , where for each  $i$  and  $\omega \in \hat{\Omega}$ ,  $\hat{\Pi}_i(\omega) = \Pi_i(\omega) \cap \hat{\Omega}$ ,  $\hat{\mathbf{s}}(\omega) = \mathbf{s}(\omega)$ , and  $\hat{b}_i(\omega) = b_i(\omega)$ . Since for each  $\omega \in \hat{\Omega}$ ,  $b_i(\omega) \subseteq \hat{\Omega}$ , it follows that  $i$  is doxastically weakly rational in the reduced belief space. Thus  $i$  is a fortiori epistemically weakly rational at  $\omega$ . This is true in each state of  $\hat{\Omega}$  and therefore epistemic weak rationality of all players is common knowledge in the reduced space. Thus, by Theorem 1,  $C$  is a non-probabilistic correlated equilibrium.

Suppose now that  $C$  is a non-probabilistic correlated equilibrium, then, by Theorem 1, there exists a knowledge space for which  $C = \mathbf{s}(\text{CKR}^{\text{w}})$ . The knowledge space can be made a belief space by defining for each  $\omega$ ,  $b_i(\omega) = \Pi_i(\omega)$ . In this belief space,  $B_i = K_i$ , and thus  $\text{CK} = \text{CB}$ .  $\square$

we can add probabilistic beliefs to a belief structure, such that  $b_i(\omega)$  is the support of the beliefs of  $i$  in  $\Pi_i(\omega)$ , and then  $B_i = B_i^1$ .

<sup>7</sup>The requirement that  $\mathbf{s}_i$  is constant on each element of  $\Pi_i$  means that  $i$  knows that she is playing the strategy she is actually playing. However, it is easy to see that even if we require that the player only *believes* that she is playing what she is actually playing, then she necessarily knows it (see, Samet, 2013, Proposition 2).

We cannot make a claim analogous to Theorem 7 for strong rationality, as the following example demonstrates.

**Example 5.** Consider the 2 by 2 game of Figure 8.

		Player 2	
		$\alpha$	$\beta$
Player 1	$\alpha$	1, 1	0, 0
	$\beta$	0, 0	0, 0

FIGURE 8. A game showing the weakness of a doxastic definition.

We consider a state space with a state  $\omega_s$  for each of the four profiles of the game, and  $s(\omega_s) = s$ . The partition  $\Pi_1$  splits the state space according to Player 1's strategy, and  $\Pi_2$  splits the state space according to Player 2's strategy. We set  $b_1(\omega_{(\beta,\alpha)}) = b_1(\omega_{(\beta,\beta)}) = \{(\beta,\beta)\}$ ,  $b_2(\omega_{(\alpha,\beta)}) = b_2(\omega_{(\beta,\beta)}) = \{(\beta,\beta)\}$ , and  $b_i(\omega) = \Pi_i(\omega)$  for all other  $i$  and  $\omega$ . Each player, when playing  $\beta$  believes that the opponent is playing  $\beta$  too. Thus a player is doxastically strongly rational when playing  $\beta$ , as  $\beta$  is not weakly dominated by  $\alpha$  relative to the opponent's strategy  $\beta$ . Obviously the players are doxastically strongly rational when they play  $\alpha$ . Thus,  $R^{sb} = \Omega$ , and it is commonly believed that the players are doxastically strongly rational in all states. Hence the profiles played in the event that there is common belief of doxastic strong rationality are all four profiles. However, the largest strong non-probabilistic correlated equilibrium consists of  $(\alpha, \alpha)$  and  $(\beta, \beta)$  only.

In order to relate doxastic strong rationality to strong non-probabilistic correlated equilibrium, we need to make an assumption about the consistency of players' beliefs, much the same consistency of beliefs is required in the characterization of correlated equilibrium by Bayesian rationality. We define and explain this consistency next.

Knowledge satisfies the truth axiom, namely, for each  $E$ ,  $K_i E \subseteq E$ . This can be written also as  $\neg K_i E \cup E = \Omega$ . The event  $\neg K_i E \cup E$  stands for the claim that either  $i$  does not know  $E$ , or  $E$ , which is equivalent to saying that if  $i$  knows  $E$ , then  $E$ . Thus, the truth axiom says that it is always true that if  $i$  knows  $E$ , then  $E$ . Belief does not satisfy the axiom of truth, namely, it is not always true that if  $i$  believes  $E$ , then  $E$ . However, it is easy to see that it is always true that  $i$  *believes* that if she believes  $E$ , then  $E$ . This claim is expressed in terms of events by  $B_i(\neg B_i E \cup E) = \Omega$ .

*Interpersonal belief consistency* holds when each player believes not only that her beliefs are true, but also that the beliefs of others are true. That is when for each  $i$  and  $j$ ,  $B_i(\neg B_j E \cup E) = \Omega$ . Obviously, the corresponding event for knowledge  $K_i(\neg K_j E \cup E) = \Omega$  trivially holds, as  $\neg K_j E \cup E = \Omega$ . Thus, consistency of knowledge is guaranteed, as whatever is known must be true. It is shown in Samet (2013) that interpersonal belief consistency holds if and only if  $\cup_{\omega \in \Omega} b_i(\omega) = \cup_{\omega \in \Omega} b_j(\omega)$  for all  $i$  and  $j$ .

**Theorem 8.** *Let  $C$  be a non-empty set of profiles. Then,  $C$  is a strong non-probabilistic correlated equilibrium if and only if  $C = \mathbf{s}(\text{CBR}^{\text{sb}})$  in some belief structure for  $G$  in which beliefs are interpersonally consistent.*

*Proof.* Suppose that a non-empty set  $C$  satisfies  $C = \mathbf{s}(\text{CBR}^{\text{sb}})$  in a belief space for  $G$  in which beliefs are interpersonally consistent. As we proved in Theorem 7, for each  $\omega \in \text{CBR}^{\text{sb}}$  and  $i$ ,  $b_i(\omega) \subseteq R_i^{\text{sb}}$  and  $b_i(\omega) \subseteq \text{CBR}^{\text{sb}}$ . We claim that For each  $i$  and  $j$ ,  $\cup_{\omega \in \text{CBR}^{\text{sb}}} b_i(\omega) = \cup_{\omega \in \text{CBR}^{\text{sb}}} b_j(\omega)$ . Indeed, let  $\omega' \in \cup_{\omega \in \text{CBR}^{\text{sb}}} b_i(\omega)$ . By interpersonal consistency there exist some  $\bar{\omega} \in \Omega$  such that  $\omega' \in b_j(\bar{\omega})$ . As  $b_j(\bar{\omega}) = b_j(\omega')$ , and since  $\omega' \in \text{CBR}^{\text{sb}}$ , it follows that  $b_j(\omega') \subseteq \text{CBR}^{\text{sb}}$ . Hence,  $\omega' \in \cup_{\omega \in \text{CBR}^{\text{sb}}} b_j(\omega)$ .

Therefore, there exists a non-empty subset  $\hat{\Omega} \subseteq \text{CBR}^{\text{sb}}$ , such that  $\hat{\Omega} = \cup_{\omega \in \text{CBR}^{\text{sb}}} b_i(\omega)$  for each  $i$ , and since for  $\omega' \in b_i(\omega)$ ,  $b_i(\omega') = b_i(\omega)$ ,  $\hat{\Omega} = \cup_{\omega \in \hat{\Omega}} b_i(\omega)$ , for each  $i$ . Thus, the family of sets  $\{b_i(\omega) \mid \omega \in \hat{\Omega}\}$  is a partition of  $\hat{\Omega}$  for each  $i$ . This makes  $(\hat{\Omega}, (b_i)_{i \in I}, \hat{\mathbf{s}})$ , where  $\hat{\mathbf{s}}$  is the restriction of  $\mathbf{s}$  to  $\hat{\Omega}$ , a knowledge structure. Since for  $\omega \in \hat{\Omega}$  and  $i$ ,  $b_i(\omega) \subseteq R_i^{\text{sb}}$ , it follows that  $\hat{\mathbf{s}}_i (= \mathbf{s}_i)$  is undominated relative to the set of  $i$ 's opponents profiles in  $b_i(\omega)$ . This means that  $i$  is epistemically strongly rational at  $\omega$  in the space  $\hat{\Omega} = \cup_{\omega \in \text{CBR}^{\text{sb}}} b_i(\omega)$ . Since this is true for each  $i$  and  $\omega \in \hat{\Omega}$ , it follows that epistemic strong rationality is common knowledge in this structure. Hence, by Theorem 1, the set of profiles played in  $\hat{\Omega}$  is a strong non-probabilistic correlated equilibrium. But this set is the same as the set of profiles played in  $\text{CBR}^{\text{sb}}$ .

If  $C$  is a strong non-probabilistic correlated equilibrium, then, by Theorem 1  $C = \mathbf{s}(\text{CKR}^{\text{s}})$  for some knowledge structure. This structure can be considered as a belief structure with  $B_i = K_i$ . Thus,  $C = \mathbf{s}(\text{CBR}^{\text{sb}})$ . Moreover, as  $B_i$  is a knowledge operator, it follows that interpersonal consistency holds in this space.  $\square$

## 7. FINAL COMMENTS

The theory of non-Bayesian rationality presented here fills a gap between Aumann (1987) and Aumann (1995, 1998). While the first paper presents a model for Bayesian rationality in games in strategic form, the other two study non Bayesian rationality in games in extensive form. This paper provides the missing link of non-Bayesian rationality in games in strategic form which characterizes non-probabilistic correlated equilibria. We have shown how the notions of substantive and material rationality can be fully characterized and expressed in terms of non-probabilistic correlated equilibria, and even reduced to it.

Our results simply and extend the theory of rationalizability. The set of strategies that survive iterated elimination of strongly dominated strategies was characterized by Tan and Werlang (1988) in terms of common knowledge of Bayesian rationality when there is no common prior. Here, in Theorem 3, we characterize this set in terms of the notion of weak rationality which is simpler and weaker than Bayesian rationality. Moreover, the analysis in Tan and Werlang (1988) is only local, while we provide both local and global analysis of the implications of common knowledge or rationality. Tan and Werlang (1988) characterized the *profiles* that can be played *locally*, that is, in a *state* where common knowledge of Bayesian rationality holds. We also provide the characterization of the same profiles by saying that a profile played in a state where common knowledge of weak rationality holds must be an element of the largest weak non-probabilistic correlated equilibrium, and hence

a profile that survives the iterated elimination of strongly dominated strategies. We also give in Theorem 1 a *global* characterization of common knowledge of weak rationality that has no counterpart in Tan and Werlang (1988). That is, we characterize the *sets* of profiles that can be played in the *event* that weak rationality is common knowledge, namely, the weak non-probabilistic correlated equilibria.<sup>8</sup>

A global characterization of common knowledge of Bayesian rationality, which was not addressed in Tan and Werlang (1988), results in exactly the same solution that is characterized by common knowledge of weak rationality: The family of the sets of profiles that can be played in the event that Bayesian rationality is common knowledge is the family of weak non-probabilistic correlated equilibria. Since we did not describe the formal model of probabilistic beliefs we don't prove the claim formally but sketch it briefly. Obviously, the set of profiles that are played in the event that Bayesian rationality is common knowledge is a non-probabilistic correlated equilibrium, since Bayesian rationality implies weak rationality. For the other direction we show that a knowledge structure can be augmented with probabilistic beliefs in such a way that when a player is weakly rational, she is Bayesian rational according to these beliefs. This is a simple separation argument such as has been used in many places in the literature. We can conclude that the version of rationalizability studied in Tan and Werlang (1988) requires only the notion of weak rationality rather than the strong notion of Bayesian rationality, for both the local and global perspectives.

#### REFERENCES

- ROBERT J. AUMANN (1974): "Subjectivity and Correlation in Randomized Strategies," *Journal of Mathematical Economics*, 1, 67–96.
- ROBERT J. AUMANN (1987): "Correlated Equilibria as an Expression of Bayesian Rationality," *Econometrica*, 55(1), 1–18.
- ROBERT J. AUMANN (1995): "Backwards Induction and Common Knowledge of Rationality," *Games and Economic Behavior*, 8(1), 6–19.
- ROBERT J. AUMANN (1998): "On the Centipede Game," *Games and Economic Behavior*, 23(1), 97–105.
- ROBERT J. AUMANN AND JACQUES H. DREZE (2008): "Rational Expectations in Games," *The American Economic Review*, 98(1), 72–86.
- B. DOUGLAS BERNHEIM (1984): "Rationalizable Strategic Behavior," *Econometrica*, 52(4), 1007–1028.
- GIACOMO BONANNO AND KLAUS NEHRING (1998): "Assessing the Truth Axiom under Incomplete Information," *Mathematical Social Sciences*, 36(1), 3–29.
- ADAM BRANDENBURGER AND EDDIE DEKEL (1987): "Rationalizability and Correlated Equilibria," *Econometrica*, 55(6), 1391–1402.
- ADAM BRANDENBURGER, AMANDA FRIEDENBERG, AND H. JEROME KEISLER (2008): "Admissibility in Games1," *Econometrica*, 76(2), 307–352.
- ALFREDO DI TILLIO, JOSEPH HALPERN, AND DOV SAMET (2012): "Conditional Belief Types," a manuscript.
- YOSSI FEINBERG (2000): "Characterizing Common Priors in the Form of Posteriors," *Journal of Economic Theory*, 91(2), 127–179.

<sup>8</sup>Aumann (1987) is a global characterization of common knowledge of Bayesian rationality when there is a common prior. Aumann and Dreze (2008) give a local characterization of the same condition.



- FRANOISE FORGES (1990): "Correlated Equilibrium in Two-Person Zero-Sum Games," *Econometrica*, 58(2), 515.
- DOV MONDERER AND DOV SAMET (1989): "Approximating Common Knowledge with Common Beliefs," *Games and Economic Behavior*, 1(2), 170–190.
- STEPHEN MORRIS (1994): "Trade with Heterogeneous Prior Beliefs and Asymmetric Information," *Econometrica*, 62(6), 1327–1347.
- ROGER MYERSON (1991): *Game Theory: Analysis of Conflict*, Harvard University Press, Cambridge, MA.
- DAVID G. PEARCE (1984): "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica*, 52(4), 1029–1050.
- DOV SAMET (1996): "Hypothetical Knowledge and Games with Perfect Information," *Games and Economic Behavior*, 17(2), 230–251.
- DOV SAMET (1998a): "Common Priors and Separation of Convex Sets," *Games and Economic Behavior*, 24(1–2), 172–174.
- DOV SAMET (1998b): "Iterated Expectations and Common Priors," *Games and Economic Behavior*, 24(1–2), 131–141.
- DOV SAMET (2011): "On the Dispensable Role of Time in Games of Perfect Information," a manuscript.
- DOV SAMET (2013): "Common Belief of Rationality in Games of Perfect Information," *Games and Economic Behavior*, 79, 192–200.
- TOMMY CHIN-CHIU TAN AND SÉRGIO RIBEIRO DA COSTA WERLANG (1988): "The Bayesian Foundations of Solution Concepts of Games," *Journal of Economic Theory*, 45(2), 370–391.