INNOCENZO GASPARINI INSTITUTE FOR ECONOMIC RESEARCH

Institutional Members: CEPR, NBER and Università Bocconi

# WORKING PAPER SERIES

## Context-Dependent Forward Induction Reasoning

*Pierpaolo Battigalli and Amanda Friedenberg*

**Working Paper n. 351**

**This Version: March, 2012**

IGIER – Università Bocconi, Via Guglielmo Röntgen 1, 20136 Milano –Italy
http://www.igier.unibocconi.it

# Context-Dependent Forward Induction Reasoning*

Pierpaolo Battigalli† Amanda Friedenberg‡

**Abstract**

This paper studies the case where a game is played in a particular context. The context influences what beliefs players hold. As such, it may affect forward induction reasoning: If players rule out specific beliefs, they may not be able to rationalize observed behavior. The effects are not obvious. Context-laden forward induction may allow outcomes precluded by context-free forward induction. At the formal level, forward induction and contextual reasoning are defined within an epistemic structure. In particular, we represent contextual forward induction reasoning as "rationality and common strong belief of rationality" (RCSBR) within an arbitrary type structure. (The concept is due to Battigalli-Siniscalchi [8, 2002].) We ask: What strategies are consistent with RCSBR (across all type structures)? We show that the RCSBR is characterized by a solution concept we call Extensive Form Best Response Sets (EFBRS's). We go on to study the EFBRS concept in games of interest.

Forward induction is a basic concept in game theory. It reflects the idea that players rationalize their opponents' behavior, whenever possible. In particular, players form an assessment about the future play of the game, given the information about the past play and the presumption that their opponents are strategic. This affects the players' choices.

Here, we study the implications of forward induction reasoning when there is a context to the game. Because there is such a context, certain beliefs may be ruled out, and this may limit the ability of players to rationalize past behavior. As such, the context may affect forward induction reasoning.

Take the following illustrative example: It is transparent that all players think that "players all drive on the right side of the road, irrespective of whether they are driving north or south." Suppose,

*We are indebted to Adam Brandenburger, John Nachbar, and Marciano Siniscalchi for many helpful conversations. Jeff Ely and three referees provided important input—much thanks. We also thank Ethan Bueno de Mesquita, Alfredo Di Tillio, Alejandro Manelli, Elena Manzoni, Andres Perea, Larry Samuelson, Adam Szeidl, and seminar participants at Bocconi University, Boston University, Maastricht University, New York University, Northwestern University, Paris School of Economics, Toulouse, Stony Brook, UC Berkeley, the 2009 Southwest Economic Theory Conference, the 2009 North American Econometric Society Meetings, the Kansas Economic Theory Conference, the 2009 SAET conference, and the 2009 European Econometric Society Meetings for important input. Battigalli thanks MIUR and Bocconi University. Friedenberg thanks the W.P. Carey School of Business and the Olin Business School. cdfiNewWP-11-20-10.rap

†Department of Economics and IGIER, Bocconi University, 1 Via Roentgen, 20136 Milan (Italy), pierpaolo.battigalli@unibocconi.it, http://www.igier.unibocconi.it/battigalli.

‡Department of Economics, Arizona State University, W.P. Carey School of Business, P.O. Box 873806, Tempe, AZ 85287-3806, amanda.friedenberg@asu.edu, http://www.public.asu.edu/~afrieden/.

further, that it is transparent that players don't like automobile accidents. Then, if Ann actually sees Bob drive on the left side of the road, she cannot justify his past behavior. In particular, she cannot justify his behavior by maintaining a hypothesis that Bob thinks she will drive on the left side of the road—after all, it is transparent that Ann will drive on the right side of the road.

To formalize the notion of context-dependent forward induction reasoning, we need some epistemic apparatus: We need to specify what beliefs players do vs. do not consider possible, and the epistemic structure allows just that. Within the epistemic structure, we analyze forward induction reasoning. The formalization of forward induction rests on Battigalli-Siniscalchi's [8, 2002] "strong belief" idea. (See also Stalnaker [29, 1998].)

We ask: Can we characterize the strategies consistent with context-dependent forward induction reasoning? That is, can we identify the play consistent with context-dependent forward induction reasoning, without actually specifying the particular epistemic structure? Indeed we can. We show that context-dependent forward induction reasoning is captured by a solution concept we call an **extensive-form best response set** (**EFBRS**). In general, there may be many EFBRS's for a given game. Which EFBRS obtains depends on the given context.

While the EFBRS definition is new, we will see that it is equivalent to one already proposed in the literature—namely, the Directed Rationalizability concept. This solution concept is due to Battigalli-Siniscalchi [9, 2003], who refer to it as $\Delta$-rationalizability. We will discuss the connection below.

The paper proceeds as follows. We begin, in Section 1, with a heuristic treatment. This gives an overview of the concepts in the paper, and explains why the EFBRS concept captures context-dependent forward induction reasoning. It also explains the connection to Directed Rationalizability. We then turn to the formal treatment. The game and epistemic structure are defined in Sections 2-3. Rationality and strong belief are defined in Section 4. Section 5 gives the main theorem. We then turn to applications, in Sections 6-7. Finally, in Section 8, we conclude by discussing certain conceptual and technical aspects of the paper.

# 1 Heuristic Treatment

Consider the game of Battle of the Sexes (BoS) with an Outside Option, as given in Figure 1.1. The standard forward induction analysis results in Bob playing *In-Right* and Ann playing *Down*: Begin with the observation that, independent of Bob's belief, the strategy *In-Left* cannot be rational (for Bob). In particular, the strategy *Out* dominates *In-Left* at the beginning of the tree.[1] But, notice, the strategy *In-Right* may very well be consistent with rationality, e.g., if Bob assigns probability one to Ann playing *Down*, then *In-Right* is a sequential best response. If this is indeed the case, then conditional upon Ann's information set being reached, she should rationalize Bob's past behavior, assigning probability one to Bob playing *In-Right*. With this, Ann should play *Down*. Now, if

[1] Note, we often conflate a strategy with its associated plan of action. No confusion should result.

Bob begins the game understanding that Ann is rational and rationalizes past behavior, Bob should begin the game assigning probability one to *Down*. In this case, Bob should indeed play *In-Right*.

This is the standard forward induction analysis—in the spirit of Kohlberg-Mertens [19, 1986]. (See, Hillas-Kohlberg [17, 2002; Section 11].) But, arguably, this is an incomplete understanding of forward induction.

To see this, consider the case where society has formed a "lady's choice convention." Loosely: Everyone in society thinks that, if the lady gets to move in a BoS-like situation, she makes choices that can lead to her "best payoff." And, moreover, it is "transparent" that everyone thinks this. Let us ask, in this case, what are the implications of forward induction reasoning? And, when there is such a convention, might the lady, perhaps, behave in a manner consistent with the convention (in this game)?

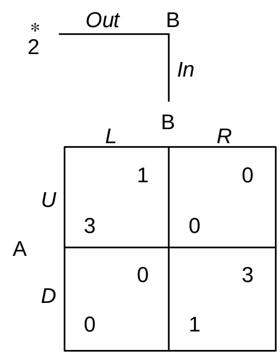


Figure 1.1

Because there is a lady's choice convention, it is thought that, if Ann gets to move, she will play *Up*, hoping to get the outcome of 3. Therefore, a rational Bob plays *Out*. Now, if Ann is given the opportunity to move, she can no longer rationalize Bob's behavior—after all, it is transparent that Bob believes she will play *Up* and, given this, a rational Bob should have played *Out*. As such, she must maintain the hypothesis that Bob is irrational. In this case, conditional upon her information set being reached, she may very well think that Bob is playing the irrational strategy *In-Left*. If she does, *Up* is indeed a best response. So, if Ann is afforded the opportunity to move, she may very well make the choice that allows her "best payoff."

Thus, *Out* is consistent with forward induction reasoning, under the convention. Of course, the argument we gave is informal. Can it be formalized? This is what we turn to next.

## 1.1 The Epistemic Game

We begin by formalizing the idea that a certain event may be transparent to the players. To do so, we append to the game an epistemic type structure. Let us review the basic elements.

There are two ingredients of an epistemic type structure: First, for each player, there are type sets $T_a$ and $T_b$. Informally, each player "knows" his own type, but faces uncertainty about the strategy the other player will choose and the type of the other player. So, each type $t_a \in T_a$ is associated with a belief on $S_b \times T_b$. Of course, we want to specify a belief at each information set. Therefore, we map each type into a conditional probability system (CPS) on $S_b \times T_b$, where the conditioning events correspond to the information sets in the game-tree. That is, for each type, there is an array of probability measures on $S_b \times T_b$, one for each information set, and this array satisfies the rules of conditional probability when possible. We write $\beta_a$ for the map from $T_a$ to CPS's on $S_b \times T_b$, and, likewise, with $a$ and $b$ interchanged.

How would we model the case of a lady's choice convention (as applied to the game in Figure 1.1)? We will have type sets $T_a$ and $T_b$. Ann's beliefs will be captured by CPS's on $S_b \times T_b$. In particular, each type of Ann will be mapped to a CPS on $S_b \times T_b$. Specifically, for each such CPS, there will be a type of Ann, viz. $t_a$, so that $\beta_a(t_a)$ is exactly that CPS. Likewise, Bob's beliefs will be captured by CPS's on $S_a \times T_a$. In particular, now, each type of Bob will be mapped to a CPS on $S_a \times T_a$ that assigns probability one to $\{Up\} \times T_a$ at each information set. Specifically, for each such CPS, there will be a type of Bob, viz. $t_b$, so that $\beta_b(t_b)$ is exactly that CPS. (Such a structure exists. See Appendix A)

Why do these conditions capture the lady's choice convention? Note, at each information set, each type of Bob assigns probability one to the event "Ann plays *Up*," i.e., to Ann trying to achieve her "best payoff." Likewise, at each information set, each type of Ann assigns probability one to the event "at each information set, Bob assigns probability one to Ann's playing *Up*." And so on. In this sense, it is "transparent" that Bob believes that, if Ann gets to move, she will play *Up*. Appendix A formalizes the idea that an event is "transparent."

Note, the context of the strategic situation determines which beliefs are (or are not) part of the type structure. Thus, the epistemic type structure is part of the description of the strategic situation. Put differently, the strategic situation is described by a game (i.e., a game form and payoff functions) plus an epistemic type structure. We call this the epistemic game.

## 1.2 Forward Induction Reasoning

Now, to formalize the idea of forward induction reasoning: Under an epistemic analysis, we talk about a type of Ann "rationalizing" Bob's past behavior, when possible. We ask that a type of Ann maintain a hypothesis that Bob is rational, provided the information she has learned is consistent with this event. In this case, we say that the type of Ann **strongly believes** the event "Bob is rational." (The idea of strong belief is due to Battigalli-Siniscalchi [8, 2002].) Of course, we will ask for more—we will ask that Ann strongly believes the event "Bob is rational and Bob strongly believes I am rational," etc...

Return to Figure 1.1 and append to the game the epistemic type structure described in Section 1.1. Let us understand forward induction reasoning within this structure.

Begin with rationality. This is a property of a strategy-type pair, i.e., $(s_a, t_a)$ is **rational** if $s_a$ is sequentially optimal under the CPS $\beta_a(t_a)$. In our example, there are rational strategy-type pairs $(s_a, t_a)$, where $s_a$ is *Up*. There are also rational strategy-type pairs $(s_a, t_a)$, where $s_a$ is *Down*. Turn to Bob. Here, each type $t_b$ assigns probability one to *Up* (at each information set). So, the set of rational strategy-type pairs for Bob is $\{Out\} \times T_b$.

So we have: If each player is rational at $(s_a, t_a, s_b, t_b)$, then Bob plays *Out*. But, is such a state consistent with forward induction reasoning? To answer this, note there are types $t_a$ that begin the game by assigning probability one to the event $\{Out\} \times T_b$. As such, these types begin the game with a hypothesis that Bob is rational. If Ann's information set is reached, Bob cannot be rational. With this, any such type of Ann strongly believes that Bob is rational. So, there are strategy-type pairs $(s_a, t_a)$ that are rational and strongly believe Bob is rational. For these pairs, we can again have $s_a$ being *Up* or *Down*. Now turn to Bob. Each type of Bob assigns probability one to Ann's playing *Up*, and there are rational strategy type pairs (*Up*, $t_a$). So, we can find types of Bob that assign probability one to Ann's rationality at each information set. Certainly these types strongly believe Ann is rational. Thus, there are strategy-type pairs $(s_b, t_b)$ that are rational and strongly believe Ann is rational. For these pairs, we have that $s_b$ is *Out*.

Continuing along these lines, we get that, for each $m$, (i) there are states consistent with "rationality and $m^{th}$-order strong belief of rationality," and (ii) at any such state, Bob plays *Out* and Ann plays either *Up* or *Down*.

## 1.3 The Question

We have seen that context-dependent forward induction reasoning may result in an outcome precluded by the standard forward induction analysis. To see this, we fixed a particular type structure and analyzed RCSBR within the associated epistemic game.

More generally, given the full epistemic game, we can identify the context-dependent strategies by analyzing RCSBR. But, what if we (i.e., the analysts) are not given the full epistemic game—that is, what if we are only given the game tree? Are there observable implications of RCSBR across all contexts? Can we identify the strategies consistent with context-dependent forward induction reasoning, by looking only at the game tree? Put differently, what sets of strategies are consistent with context-dependent forward induction reasoning (across all contexts)? This is the main question we ask here.

We will characterize the strategies consistent with RCSBR (across all type structures). In particular, a set of strategies is consistent with RCSBR (in some structure) if and only if it satisfies certain properties defined on the game tree alone. This will be the basis for the extensive-best response set concept we mentioned in the Introduction. Using the properties of extensive-form best response sets, we will be able to make a connection to an old solution concept, namely Directed Rationalizability (Battigalli-Siniscalchi [9, 2003]).

## 1.4 Rationality and Common Strong Belief of Rationality

Let us begin with an arbitrary epistemic game. Refer to Figure 1.2. Here, $\mathbf{R0SBR}_a$ ($\mathbf{R0SBR}_b$) is the set of Ann's (resp. Bob's) rational strategy-type pairs. $\mathbf{R1SBR}_a$ ($\mathbf{R1SBR}_b$) is the set of Ann's strategy-type pairs that are rational and strongly believe "Bob is rational." More generally, $\mathbf{RmSBR}_a$ (resp. $\mathbf{RmSBR}_b$) is the set of strategy-type pairs for Ann (resp. Bob) that are consistent with **rationality and $m^{th}$-order strong belief of rationality**.

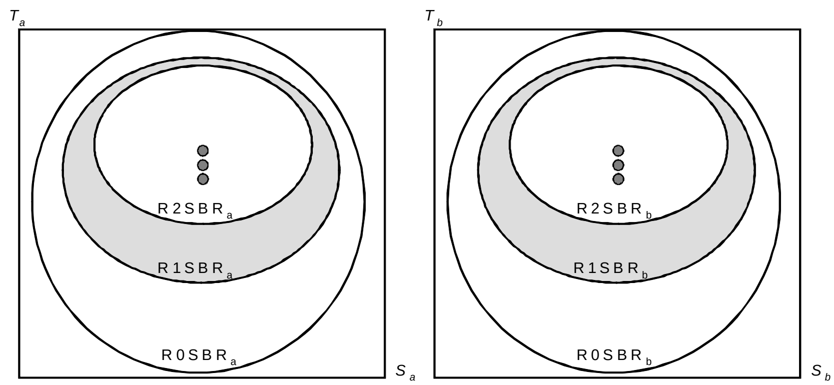


Figure 1.2

We are interested in the set of states consistent with **rationality and common strong belief of rationality** (**RCSBR**). Refer to Figure 1.3. This is the set $\mathrm{RCSBR}_a \times \mathrm{RCSBR}_b$, where $\mathrm{RCSBR}_a$ (resp. $\mathrm{RCSBR}_b$) is the intersection of the sets $\mathrm{R}m\mathrm{SBR}_a$ (resp. $\mathrm{R}m\mathrm{SBR}_b$) across all $m$. Can we characterize the strategies played under RCSBR, i.e., the set $Q_a \times Q_b$ in Figure 1.3? For this, fix some $s_a \in Q_a$ and note that there is some type $t_a$ so that $(s_a, t_a)$ is contained in $\mathrm{R}m\mathrm{SBR}_a$, for each $m$. We will use this to identify two facts about $s_a$.

For the first fact: Note that $s_a$ is optimal under the CPS associated with $t_a$, namely $\beta_a(t_a)$. It follows that $s_a$ is optimal under the marginal of $\beta_a(t_a)$ on $S_b$ (a CPS on Bob's strategies). For the second fact, note that $t_a$ strongly believes the event $\mathrm{R0SBR}_b$, the event $\mathrm{R1SBR}_b$, the event $\mathrm{R2SBR}_b$, etc. So, by a conjunction property of strong belief, $t_a$ strongly believes the event $\mathrm{RCSBR}_b$. It then follows from a marginalization property of strong belief that the marginal of $\beta_a(t_a)$ on $S_b$ strongly believes $Q_b$.

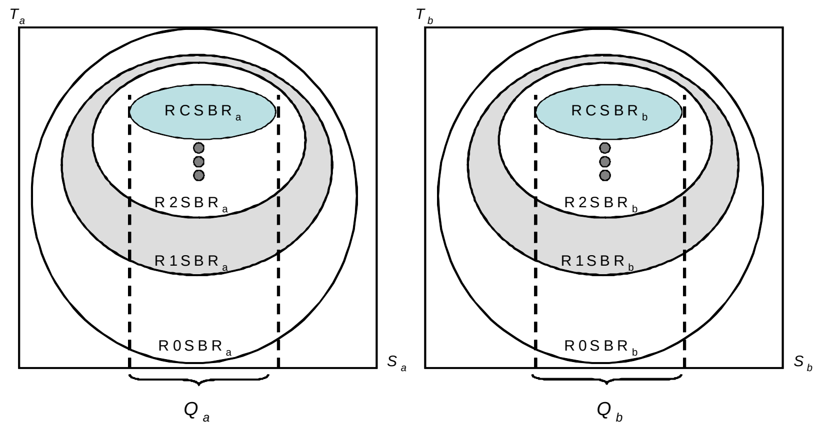


Figure 1.3

So we have:

For each $s_a \in Q_a$, there is a CPS on $S_b$, viz. $\mu_a(s_a)$, so that

(i) $s_a$ is sequentially optimal under $\mu_a(s_a)$, and

(ii) $\mu_a(s_a)$ strongly believes $Q_b$;

and likewise with $a$ and $b$ interchanged.

In sum: For a given type structure, the projection of the RCSBR set into $S_a \times S_b$ satisfies conditions (i)-(ii). But, do these conditions characterize RCSBR? In particular, given a set $Q_a \times Q_b$ satisfying conditions (i) and (ii), can we construct a type structure so that $Q_a \times Q_b$ is the projection of the RCSBR set into $S_a \times S_b$? The answer may be no.

## 1.5 Maximality

Consider the game in Figure 1.4, and the set $Q_a \times Q_b = \{Out\} \times \{Left, Center\}$. This set satisfies conditions (i)-(ii) in Section 1.4. Begin with Ann and consider the CPS that assigns probability $\frac{1}{2} : \frac{1}{2}$ to $Left$ : $Center$, at each information set. The strategy $Out$ is sequentially optimal under this CPS. Of course, this CPS strongly believes $Q_b$. Turning to Bob, consider a CPS that assigns probability one to $Out$ at the initial node and probability $\frac{1}{4} : \frac{1}{4} : \frac{1}{2}$ to $In$-$Up$ : $In$-$Middle$ : $In$-$Down$ conditional upon Bob's subgame being reached. The strategies $Left$ and $Center$ are sequentially optimal under this CPS and this CPS strongly believes $Q_a$. So, conditions (i)-(ii) are satisfied for $Q_a \times Q_b$.

Note, however, there is no type structure so that the projection of the RCSBR set into $S_a \times S_b$ is $Q_a \times Q_b$. In fact, we can go further: There is no type structure so that $Out$ is consistent with RCSBR.

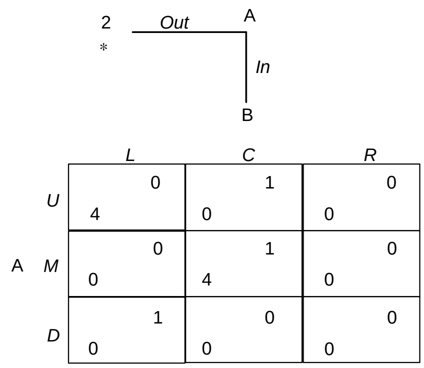


Figure 1.4

To see this, suppose otherwise, i.e., that we have found a type structure so that Ann's playing *Out* is consistent with RCSBR. Then we have a type $t_a$ so that (*Out*, $t_a$) is consistent with RCSBR. Certainly, (*Out*, $t_a$) is rational, and $t_a$ strongly believes the event "Bob is rational." Since each pair in $\{Right\} \times T_b$ is irrational and $t_a$ strongly believes "Bob is rational," the type $t_a$ is associated with a CPS that (at each node) assigns probability one to $\{Left, Center\} \times T_b$. Now, since (*Out*, $t_a$) is rational, the associated CPS must assign probability $\frac{1}{2} : \frac{1}{2}$ to $\{Left\} \times T_b : \{Center\} \times T_b$, at each node. With this, (*In-Up*, $t_a$) and (*In-Middle*, $t_a$) are also rational. Indeed, since $t_a$ strongly believes each of the R$m$SBR$_b$ events, both (*In-Up*, $t_a$) and (*In-Middle*, $t_a$) must be consistent with RCSBR.

Next, consider an RCSBR strategy-type pair for Bob, viz. $(s_b, t_b)$. Conditional upon Bob's information set being reached, $t_b$ must assign probability one to $\{In\text{-}Up, In\text{-}Middle\} \times T_a$. (To see this, note that this event contains rational strategy-type pairs, while the event $\{In\text{-}Down\} \times T_a$ does not contain any rational strategy-type pairs.) So, since $(s_b, t_b)$ is rational, $s_b = Center$. As such, the RCSBR strategy-type pairs for Bob are contained in $\{Center\} \times T_b$. But, now notice that the CPS associated with $t_a$ does not strongly believe the RCSBR event for Bob. This is a contradiction.

Let us ask: What went wrong in this example? We began with a set $Q_a \times Q_b$ satisfying conditions (i)-(ii). In particular, we had a strategy $s_a \in Q_a$ for which there was a unique CPS $\mu_a(s_a)$, so that $s_a$ and $\mu_a(s_a)$ satisfy conditions (i)-(ii). But, under this CPS, we had a strategy $r_a \in S_a \backslash Q_a$ that was also sequentially optimal. (Actually, there were two such sequentially optimal strategies in $S_a \backslash Q_a$.) As such, if $(s_a, t_a)$ is consistent with RCSBR, then $(r_a, t_a)$ must also be consistent with RCSBR. That is, $Q_a$ may exclude some strategy of Ann consistent with RCSBR. If so we may be able to find a CPS $\mu_b(s_b)$ (on $S_a$) that satisfies conditions (i)-(ii) for $s_b$, despite the fact that $s_b$ is not optimal under any CPS (on $S_a \times T_a$) that strongly believes the RCSBR strategy-type pairs for Ann.

This suggests that we need to add the following maximality criterion to conditions (i)-(ii) of Section 1.4:

(iii) If $r_a \in S_a$ is sequentially optimal under $\mu_a(s_a)$, then $r_a \in Q_a$.

We will call a set an **extensive-form best response set** (**EFBRS**) if, for each $s_a \in Q_a$ there is some CPS $\mu_a(s_a)$ satisfying conditions (i)-(ii)-(iii), and likewise with $a$ and $b$ interchanged.

## 1.6 Extensive-Form Best Response Sets

Now we are ready to state the main result, namely a characterization theorem.

**Main Theorem**

*(i)* Fix an extensive-form game and an epistemic type structure. The strategies consistent with RCSBR form an EFBRS.

*(ii)* Fix an extensive-form game and an associated EFBRS, namely $Q_a \times Q_b$. Then there exists an epistemic type structure, so that the strategies consistent with RCSBR are exactly $Q_a \times Q_b$.

Return to the Battle of the Sexes with an Outside Option. For that game, there are three EFBRS's, namely $\{Out\} \times \{Up\}$, $\{Out\} \times \{Up, Down\}$, and $\{In\text{-}Right\} \times \{Down\}$. Thus, each of these solutions are consistent with forward induction reasoning. Which set obtains depends on the context within which the game is played, i.e., depends on which events are "transparent" to the players. See Appendix A for more on this point.

Note, the EFBRS $\{In\text{-}Right\} \times \{Down\}$ corresponds to the usual forward induction analysis. One situation where this EFBRS obtains is a specific "context-free" case, where all beliefs are present. Indeed, this set is also the extensive-form rationalizable (EFR) strategy set. When the type structure contains all possible beliefs—formally, when the maps $\beta_a$ and $\beta_b$ are onto—the projection of the RCSBR set into $S_a \times S_b$ is the extensive-form rationalizable strategy set.[2] See Proposition 6 in Battigalli-Siniscalchi [8, 2002] for a formal statement.

## 1.7 Directed Rationalizability

Return to the "lady's choice convention," and the associated type structure in Section 1.1. There, each type of Bob was associated with some CPS that assigned probability one to $\{Up\} \times T_a$. This gives a restriction on Bob's first-order beliefs, i.e., his beliefs about what Ann will choose. Let $\Delta_b$ represent this restriction on first-order beliefs. So, $\Delta_b$ is a subset of the CPS's on $S_a$ and, in our example, $\Delta_b$ (only) contains the CPS which assigns probability one to $Up$. We did not have a restriction on Ann's first order beliefs. So, we will write $\Delta_a$ for the set of all CPS's on $S_b$.

[2] The condition that $\beta_a$ and $\beta_b$ are onto is known as completeness. It is due to Brandenburger [14, 2003].

With $\Delta = \Delta_a \times \Delta_b$ in hand, we can take an iterative approach to analyzing the game tree—much like a "typical rationalizability" procedure. On round one, we eliminate $In\text{-}Left$ and $In\text{-}Right$ for Bob, since these strategies are not sequentially optimal under the CPS in $\Delta_b$. We do not eliminate any of Ann's strategies, since they are each sequentially optimal under some CPS (in $\Delta_a$). So, on round one, we are left with the set $\{Out\} \times \{Up, Down\}$. On round two, we note that $Out$ is sequentially optimal under the CPS in $\Delta_b$ and that CPS strongly believes $\{Up, Down\}$. Thus, we cannot eliminate any strategy of Bob on round two. Likewise, $Up$ (resp. $Down$) is sequentially optimal under a CPS that assigns probability one to $Out$ at the initial node, and probability one to $Left$ (resp. $Right$) at Bob's subgame. This CPS is contained in $\Delta_a$ and strongly believes $\{Out\}$. So, we also get $\{Out\} \times \{Up, Down\}$ on round two. Indeed, a standard induction argument gives that $\{Out\} \times \{Up, Down\}$ is the outcome of the procedure. Of course, this was the EFBRS we identified in Section 1.2.

The procedure used above is called **$\Delta$-rationalizability**, due to Battigalli [4, 1999] and further analyzed by Battigalli-Siniscalchi [9, 2003].[3] The procedure begins by fixing a set of **first-order beliefs**, i.e., a set $\Delta = \Delta_a \times \Delta_b$, where $\Delta_a$ is a set of CPS's on $S_b$ and $\Delta_b$ is a set of CPS's on $S_a$. On round one, it eliminates any strategy of Ann (resp. Bob) that is not sequentially optimal under some CPS in $\Delta_a$ (resp. $\Delta_b$). On round two, it further eliminates any strategy of Ann (resp. Bob) that is not sequentially optimal under a CPS in $\Delta_a$ (resp. $\Delta_b$) that strongly believes the round-one strategies of Bob (resp. Ann). And so on.

Note, there may be many $\Delta$-rationalizable sets—each of which is obtained by beginning the procedure with a different set of first-order beliefs $\Delta = \Delta_a \times \Delta_b$. Collecting all the $\Delta$-rationalizable sets together, we get the solution concept of **Directed Rationalizability**. The idea is that each set of first-order beliefs $\Delta = \Delta_a \times \Delta_b$ is used to direct the rationalizability procedure.

## 1.8 An Alternate Characterization Theorem

In Section 1.7, we considered a particular context. Specifically, there was a set of first-order beliefs and these beliefs were "transparent" to the players. We then used this set of first-order beliefs to compute the associated $\Delta$-rationalizable strategy set. We got the answer $\{Out\} \times \{Up, Down\}$. It turned out that this was one of the EFBRS's we identified in Section 1.6, and so is consistent with RCSBR. More generally, beginning with any set of first-order beliefs, viz. $\Delta = \Delta_a \times \Delta_b$, we can always find an epistemic structure so that the $\Delta$-rationalizable strategy set is the set of strategies consistent with RCSBR. In particular, we will see that the $\Delta$-rationalizable strategy set forms an EFBRS—so, the claim follows from part (ii) of the Main Theorem.

But, what about a converse? In Section 1.7 we began with an epistemic structure and noted

[3] Battigalli [4, 1999] and Battigalli-Sinsicalchi [9, 2003] introduced the concept to study a different problem from the one studied here. In their problem, the set $\Delta$ is given to the analyst. In our problem, we construct the set $\Delta$. See Section 8a. We also point out that [9, 2003] use a different definition, which is equivalent to the original one due to [4, 1999] in an important special case studied in their paper. Here we cannot confine ourself to that special case and we use the definition of [4, 1999], which is the conceptually correct one (see Battigalli-Prestipino [5, 2010]).

that we can compute the RCSBR strategy set by beginning with some set of first-order beliefs, viz. $\Delta = \Delta_a \times \Delta_b$, and performing the $\Delta$-rationalizability procedure. Does this hold more generally? Beginning with some epistemic structure and the RCSBR strategy set, do we always get some $\Delta$-rationalizable set (i.e., for some set $\Delta = \Delta_a \times \Delta_b$)?

One might think that the answer is no. After all, the Directed Rationalizability begins with a set of first-order beliefs and only uses this set to direct the rationalizability procedure. This seems like the right approach in a case like the lady's choice convention, where the event that is transparent corresponds to a restriction players' first-order beliefs. Recall, there, we had a restriction on Bob's first-order beliefs and we only had a restriction on Ann's second-order beliefs in so far as Ann must think that Bob thinks she will play *Up*, etc. In the case where there is a "transparent restriction" on higher-order beliefs, perhaps we cannot restrict attention to Directed Rationalizability.

We will see that this intuition is wrong. Beginning with an epistemic structure and the RCSBR strategy set, we will get some $\Delta$-rationalizable set. But, importantly, the approach taken to find this $\Delta$-rationalizable set may be different from the approach we took in Section 1.7. This is precisely because the type structure may involve a "transparent restriction" on higher-order beliefs. Let us see this more precisely—to do so, we will mimic the route we took in Section 1.7.

In Section 1.7, we began with an epistemic structure and used the structure itself to form the set $\bar{\Delta} = \bar{\Delta}_a \times \bar{\Delta}_b$. Specifically, for each type $t_a \in T_a$, consider the marginal of $\beta_a(t_a)$ on $S_b$. These CPS's form the set $\bar{\Delta}_a$. Construct the set $\bar{\Delta}_b$ analogously. Note, here, the strategies that survive one round of $\bar{\Delta}$-rationalizability are exactly the strategies that are consistent with $\text{R0SBR}_a \times \text{R0SBR}_b$. But, on round two, we lose the equivalence: If $\beta_a(t_a)$ strongly believes the event "Bob is rational," then the marginal of $\beta_a(t_a)$ will also strongly believe that "Bob chooses a strategy consistent with one round of elimination of $\bar{\Delta}$-rationalizability." (Here, we use a marginalization property of strong belief, plus the round-one equivalence.) But, the converse need not hold. So, the strategies that survive two rounds of $\bar{\Delta}$-rationalizability may strictly contain the R1SBR strategies. And, on round three, we loose the inclusion. If the CPS $\beta_a(t_a)$ strongly believes the R1SBR event for Bob, then the marginal of $\beta_a(t_a)$ will also strongly believe that "Bob chooses a strategy consistent with R1SBR." But, recall, the strategies consistent with R1SBR may be strictly contained in the strategies that survive two rounds of $\bar{\Delta}$-rationalizability. So there may be information sets consistent with this latter event, but not the former. This implies that, even if $\beta_a(t_a)$ strongly believes the R1SBR event for Bob, it need not strongly believe that Ann's behavior is consistent with two rounds of $\bar{\Delta}$-rationalizability. As such, we loose any relationship between the RCSBR strategies and the $\bar{\Delta}$-rationalizable strategy set.

But, there is another route, that instead uses the EFBRS properties to form a set $\Delta = \Delta_a \times \Delta_b$ of first-order beliefs. Fix an epistemic structure. The RCSBR strategies form an EFBRS, viz. $Q_a \times Q_b$. For each $s_a \in Q_a$, we have some CPS $\mu_a(s_a)$ satisfying conditions (i)-(ii)-(iii) above. Take $\Delta_a$ to be the set of such CPS's, i.e., one for each $s_a \in Q_a$, and construct $\Delta_b$ similarly. Now we do have an equivalence between the RCSBR strategies and the $\Delta$-rationalizable strategies. More precisely, for each $m \geq 1$, $Q_a \times Q_b$ is the set of strategies that survives $m$-rounds of elimination of $\Delta$-rationalizability. The case of $m = 1$ follows from properties (i) and (iii) of an EFBRS. The case of $m = 2$ uses condition (ii) of an EFBRS. And so on, by induction.

In sum:

**Alternate Characterization Theorem**

*(i)* Fix an extensive-form game and an epistemic type structure. There exists a set of first-order beliefs $\Delta = \Delta_a \times \Delta_b$ so that the set of strategies consistent with RCSBR is exactly the $\Delta$-rationalizable strategy set.

*(ii)* Fix an extensive-form game and a set of first-order beliefs $\Delta = \Delta_a \times \Delta_b$. Then there exists an epistemic type structure, so that the set of strategies consistent with RCSBR is exactly the $\Delta$ -rationalizable strategy set.

Indeed, the solution concept of Directed Rationalizability characterizes RCSBR. The different components of this solution correspond to different sets of first-order beliefs. But, note, it may not be obvious which $\Delta$-rationalizable set is associated with a particular type structure. To see this, fix a type structure and let $\bar{\Delta}_a \times \bar{\Delta}_b$ be the set of all first-order beliefs associated with that particular structure. The RCSBR strategies form some $(\Delta_a \times \Delta_b)$-rationalizable set, but this set may be distinct from the $\left(\bar{\Delta}_a \times \bar{\Delta}_b\right)$-rationalizable set.

To repeat: While, in general, it may not be obvious which $\Delta$-rationalizable set may be associated with a particular type structure, there is one important case where there is an obvious connection: This is the case where, in a certain sense, the only restriction on players' beliefs amounts to a restriction on first-order beliefs and its transparency, as in the "lady's choice convention" discussed above. Battigalli-Prestipino [5, 2010] provide a formal statement. They show that, in this case, the RCSBR strategy set does correspond to the $\left(\bar{\Delta}_a \times \bar{\Delta}_b\right)$-rationalizable strategy set. (See Section 8a below.)

## 1.9 Analyzing Games: The EFBRS Properties

We have seen that the Directed Rationalizability solution concept also characterizes RCSBR. To show this, we show it is equivalent to the EFBRS concept. In particular, we begin with the RCSBR strategies $Q_a \times Q_b$. We make use of the fact that $Q_a \times Q_b$ satisfies the EFBRS properties to show that we can find some set of first-order beliefs, viz. $\Delta = \Delta_a \times \Delta_b$, so that the $\Delta$-rationalizable set is $Q_a \times Q_b$.

While the EFBRS and Directed Rationalizability concepts are equivalent, it will often be useful to focus on the former definition. The reason is that properties (i), (ii), and (iii) of an EFBRS give some immediate implications in terms of behavior. In Sections 6-7, we will discuss the consequences of context-dependent forward reasoning for some specific games. There, the EFBRS properties will play an important role, much in the same way that the properties of a self-admissible set

(Brandenburger-Friedenberg-Keisler [16, 2008]) play an important role in analyzing games. Indeed, we will see that these properties help to analyze games such as centipede, the finitely repeated prisoner's dilemma, and perfect information games.

In Section 8c, we return to further discuss the EFBRS vs. Directed Rationalizability definitions.

# 2 The Game

We consider finite extensive form games of perfect recall. We write $\Gamma$ for such a game. The definition we consider is similar to that in Osborne-Rubinstein [22, 1994; Definition 200.1]. In particular, it allows for simultaneous moves.[4]

There are two players, namely $a$ (Ann) and $b$ (Bob).[5] Let $C_a$ and $C_b$ be **choice** or **action sets** for Ann and Bob. A history for the game consists of (possibly empty) sequences of simultaneous choices for Ann and Bob. More formally, a **history** is either (i) the empty sequence, written $\phi$, or (ii) a sequence of choice pairs $\left(c^1, \ldots, c^K\right)$, where each $c^k = (c_a^k, c_b^k) \in C_a \times C_b$. Note, histories have the property that, if $\left(c^1, \ldots, c^K\right)$ is a history then so is $\left(c^1, \ldots, c^L\right)$, for each $L \leq K$. Note that each history can be viewed as a node in the tree. As such, we will interchangeably use the terms "node" and "history."

Write $x$ for a history of the game and let $C(x) = \{c \in C_a \times C_b : (x, c) \text{ is a history for the game}\}$. Write $C_a(x) = \text{proj}_{C_a} C(x)$ and $C_b(x) = \text{proj}_{C_b} C(x)$. By assumption, these sets have the property that $C(x) = C_a(x) \times C_b(x)$. The interpretation is that $C_a(x)$ is the set of **choices available to** $a$ **at history** $x$. If $|C_a(x)| \geq 2$, say $a$ **moves at history** $x$ or $a$ **is active at** $x$. (If $|C_a(x)| \leq 1$, $a$ is inactive at history $x$.) Call $x$ a **terminal history** of the game if $C(x) = \emptyset$. (Terminal histories can be viewed either as **terminal nodes** or **paths** for the game.)

Let $H_a$ (resp. $H_b$) be a partition of the set of all nodes at which $a$ (resp. $b$) is active plus the initial node $\phi$. The partition $H_a$ (resp. $H_b$) has the property that if $x$, $x'$ are contained in the same partition member, viz. $h$ in $H_a$ (resp. $H_b$), then $C_a(x) = C_a(x')$ (resp. $C_b(x) = C_b(x')$). The interpretation is that $H_a$ (resp. $H_b$) is the family of **information sets** for $a$ (resp. $b$). (Note that $\{\phi\} \in H_a \cap H_b$, perfect recall imposes further requirements on $H_a$ and $H_b$. See Osborne-Rubinstein [22, 1994; Definition 203.3].) Write $H = H_a \cup H_b$.

Write $Z$ for the set of terminal histories of the game, and let $z$ be an arbitrary element of $Z$. **Extensive-form payoff functions** are given by $\Pi_a : Z \to \mathbb{R}$ and $\Pi_b : Z \to \mathbb{R}$.

We abuse notation and write $C_a(h)$ for the set of choices available to $a$ at information set $h \in H_a$. With this, the set of **strategies** for player $a$ is given by $S_a = \prod_{h \in H_a} C_a(h)$. Define $S_b$ analogously. Each pair of strategies $(s_a, s_b)$ induces a path through the tree. Let $\zeta : S_a \times S_b \to Z$ map each strategy profile into the induced path. **Strategic-form payoff functions** are given by $\pi_a = \Pi_a \circ \zeta$ and $\pi_b = \Pi_b \circ \zeta$. Given a profile $(s_a, s_b)$, write $\pi(s_a, s_b) = (\pi_a(s_a, s_b), \pi_b(s_a, s_b))$ and refer to this

[4] This definition incorporates repeated games. Our analysis does not depend on the specific definition used.

[5] The analysis extends to $n$-player games, up to issues of correlation. See Section 8f.

payoff vector as an **outcome** of the game. Two strategy profiles, $(s_a, s_b)$ and $(r_a, r_b)$, are **outcome equivalent** if $\pi(s_a, s_b) = \pi(r_a, r_b)$. (Of course, if $(s_a, s_b)$ and $(r_a, r_b)$ induce the same path (i.e., if $\zeta(s_a, s_b) = \zeta(r_a, r_b)$), they are outcome equivalent. But, they may be outcome equivalent even if they do not.)

For each information set $h \in H$, write $S_a(h)$ (resp. $S_b(h)$) for the set of strategies for $a$ (resp. $b$) that allow $h$. (That is, $s_a \in S_a(h)$ if there is some $s_b \in S_b$ so that the path induced by $(s_a, s_b)$ passes through $h$.) Let $\mathcal{S}_a$ (resp. $\mathcal{S}_b$) be the collection of all $S_a(h)$ (resp. $S_b(h)$) for $h \in H_b$ (resp. $h \in H_a$). Thus, $\mathcal{S}_a$ represents the information structure of $b$ about the strategy of $a$. In particular, at each of $b$'s information sets, he will have a belief about $a$ that assigns probability one to the set of $a$'s strategies consistent with the information set being reached.

# 3 The Type Structure

This section appends to the game a type structure, within which the terms 'rationality' and 'strong belief' can be defined. Again, this section closely follows Battigalli-Siniscalchi [8, 2002].

Throughout, let $\Omega$ be a separable metrizable space and let $\mathcal{B}(\Omega)$ the Borel $\sigma$-algebra on $\Omega$. We endow the product of separable metrizable spaces with the product topology, and a subset of a separable metrizable space with the relative topology. Write $\mathcal{P}(\Omega)$ for the set of Borel probability measures on $\Omega$, and endow $\mathcal{P}(\Omega)$ with the topology of weak convergence.

**Definition 3.1 (Renyi [26, 1955])** *Fix a separable metrizable space $\Omega$ and a non-empty collection of events $\mathcal{E} \subseteq \mathcal{B}(\Omega)$. A **conditional probability system (CPS)** on $(\Omega, \mathcal{E})$ is a mapping $\mu(\cdot|\cdot) : \mathcal{B}(\Omega) \times \mathcal{E} \to [0,1]$ such that, for any $E \in \mathcal{B}(\Omega)$ and $F, G \in \mathcal{E}$,*

*(i)* $\mu(F|F) = 1$,

*(ii)* $\mu(\cdot|F) \in \mathcal{P}(\Omega)$, *and*

*(iii)* $E \subseteq F \subseteq G$ *implies* $\mu(E|G) = \mu(E|F)\,\mu(F|G)$.

*Call $\mathcal{E}$, with $\emptyset \neq \mathcal{E} \subseteq \mathcal{B}(\Omega)$, a **collection of conditioning events** for $\Omega$.*

When it is clear that $\mu(\cdot|\cdot)$ is a CPS on $(\Omega, \mathcal{E})$, we omit reference to its arguments simply writing $\mu$ instead of $\mu(\cdot|\cdot)$.

Write $\mathcal{C}(\Omega, \mathcal{E})$ for the set of conditional probability systems on $(\Omega, \mathcal{E})$. Note, $\mathcal{C}(\Omega, \mathcal{E})$ can be viewed as a subset of $[\mathcal{P}(\Omega)]^{\mathcal{E}}$. We endow $[\mathcal{P}(\Omega)]^{\mathcal{E}}$ with the product topology and, then, $\mathcal{C}(\Omega, \mathcal{E})$ with the relative topology. When $\mathcal{E}$ is a countable, $\mathcal{C}(\Omega, \mathcal{E})$ is separable metrizable. When it is clear from the context what the set of conditioning events are, we omit reference to $\mathcal{E}$, simply writing $\mathcal{C}(\Omega)$.

We will often be interested in product sets. We adopt the convention that if $\Omega_1 \times \Omega_2 = \emptyset$ then both $\Omega_1 = \emptyset$ and $\Omega_2 = \emptyset$. Fix some $\mathcal{E} \subseteq \mathcal{B}(\Omega_1)$, and write $\mathcal{E} \otimes \Omega_2$ for the set of all $E \times \Omega_2$ where $E \in \mathcal{E}$. Note that $\mathcal{E} \otimes \Omega_2 \subseteq \mathcal{B}(\Omega_1 \times \Omega_2)$.

Consider a CPS $\mu(\cdot|\cdot)$ on $(\Omega_1 \times \Omega_2, \mathcal{E} \otimes \Omega_2)$, where $\mathcal{E} \subseteq \mathcal{B}(\Omega_1)$. Define $\nu(\cdot|\cdot) : \mathcal{B}(\Omega_1) \times \mathcal{E} \to [0,1]$ so that $\nu(E|F) = \mu(E \times \Omega_2 | F \times \Omega_2)$ for all $E \in \mathcal{B}(\Omega_1)$ and $F \in \mathcal{E}$. Then $\nu$ is a conditional probability system on $(\Omega_1, \mathcal{E})$. When $\nu(\cdot|\cdot)$ is defined in this way, write $\nu(\cdot|\cdot) = \mathrm{marg}_{\Omega_1} \mu(\cdot|\cdot)$. No confusion should result.

**Definition 3.2** *Fix an extensive-form game* $\Gamma$. *A* $\Gamma$***-based type structure*** *is a collection*

$$\langle S_a, S_b; \mathcal{S}_a, \mathcal{S}_b; T_a, T_b; \beta_a, \beta_b \rangle ,$$

*where* $T_a$ *(resp.* $T_b$*) is a nonempty separable metrizable space and* $\beta_a : T_a \to \mathcal{C}(S_b \times T_b)$ *(resp.* $\beta_b : T_b \to \mathcal{C}(S_a \times T_a)$*) is a measurable belief map associated with conditioning events* $\mathcal{S}_b \otimes T_b$ *(resp.* $\mathcal{S}_a \otimes T_a$*). Members of* $T_a$ *(resp.* $T_b$*) are called* ***types****. Members of* $S_a \times T_a \times S_b \times T_b$ *are called* ***states****.*

In Section 1.1 we argued that the type structure captures the idea that certain beliefs are "transparent" to the players. This is true in a precise sense: Begin with Battigalli-Siniscalchi's [7, 1999] canonical construction of a type structure that contains all hierarchies of conditional beliefs (satisfying coherency and common belief of coherency).[6] Lets us look at the so-called self-evident events within this structure. Loosely, these are events $E \in \mathcal{B}(S_a \times T_a \times S_b \times T_b)$, where $E$ obtains and, at each information set, each player assigns probability one to $E$, each player assigns probability one to the other player assigning probability one to $E$, etc. (Appendix A provides a formal definition.) Each type structure can be mapped into the canonical construction and, in a certain sense, each type structure forms a self-evident event in the canonical construction, i.e., under this mapping. (Note, this assumes a certain bimeasurability condition.) Furthermore, each self-evident event in the canonical type structure corresponds to a "smaller" type structure. The formal treatment is provided in Appendix A.

# 4 Rationality and Strong Belief

We now turn to the main epistemic definitions, all of which have counterparts with $a$ and $b$ reversed. Begin by extending $\pi_a(\cdot,\cdot)$ to $S_a \times \mathcal{P}(S_b)$ in the usual way, i.e., $\pi_a(s_a, \varpi_a) = \sum_{s_b \in S_b} \pi_a(s_a, s_b)\, \varpi_a(s_b)$. (Notice, the measure $\varpi_a$ on $S_b$ reflects a belief by $a$ about $b$, so we write $\varpi_a \in \mathcal{P}(S_b)$.)

**Definition 4.1** *Fix* $X_a \subseteq S_a$ *and* $s_a \in X_a$*. Say* $s_a$ *is* ***optimal under*** $\varpi_a \in \mathcal{P}(S_b)$ ***given*** $X_a$ *if* $\pi_a(s_a, \varpi_a) \geq \pi_a(r_a, \varpi_a)$ *for all* $r_a \in X_a$*.*

**Definition 4.2** *Say* $s_a \in S_a$ *is* ***sequentially optimal under*** $\mu_a(\cdot|\cdot) : \mathcal{B}(S_b) \times \mathcal{S}_b \to [0,1]$ *if, for all* $h$ *with* $s_a \in S_a(h)$*,* $s_a$ *is optimal under* $\mu_a(\cdot|S_b(h))$ *given* $S_a(h)$*. Say* $s_a \in S_a$ *is* ***sequentially***

[6] Note, Battigalli-Siniscalchi's [7, 1999] canonical construction is a type structure in the sense of Definition 3.2. Specifically, in the case of a game tree, the basic conditioning events are clopen and so [7, 1999] get $T_a$ and $T_b$ to be Polish, as an output. Here, we don't require Polishness. Allowing a more general definition of a type structure simplifies some of the technical arguments.

***justifiable*** *if there exists* $\mu_a(\cdot|\cdot) : \mathcal{B}(S_b) \times \mathcal{S}_b \to [0,1]$ *so that* $s_a$ *is sequentially optimal under* $\mu_a(\cdot|\cdot)$.

**Definition 4.3** *Say* $(s_a, t_a)$ *is* ***rational*** *if* $s_a$ *is sequentially optimal under* $\text{marg}_{S_b} \beta_a(t_a)$.

Let $R_a$ be the set of strategy-type pairs, viz. $(s_a, t_a)$, at which $a$ is rational.

**Definition 4.4 (Battigalli-Siniscalchi [8, 2002])** *Fix a CPS* $\mu(\cdot|\cdot) : \mathcal{B}(\Omega) \times \mathcal{E} \to [0,1]$ *and an event* $E \in \mathcal{B}(\Omega)$. *Say* $\mu$ ***strongly believes*** $E$ *if*

*(i) there exists* $F \in \mathcal{E}$ *so that* $E \cap F \neq \emptyset$, *and*

*(ii) for each* $F \in \mathcal{E}$, $E \cap F \neq \emptyset$ *implies* $\mu(E|F) = 1$.

Fix a CPS $\mu$ that strongly believes $\dot{E}$. Note, if $\Omega \in \mathcal{E}$, then $\mu(E|\Omega) = 1$. (In our application, we will, in general, have $\Omega \in \mathcal{E}$.) Now, we point out two general properties about strong belief.

**Property 4.1 (Conjunction)** *Fix a CPS on* $(\Omega, \mathcal{E})$, *viz.* $\mu$, *and a finite or countable collection of events* $E_1, E_2, \ldots$. *If* $\mu$ *strongly believes* $E_1, E_2, \ldots$ *then* $\mu$ *strongly believes* $\bigcap_m E_m$.

**Property 4.2 (Marginalization)** *Fix a CPS* $\mu$ *on* $(\Omega_1 \times \Omega_2, \mathcal{E} \otimes \Omega_2)$, *where* $\emptyset \neq \mathcal{E} \subseteq \mathcal{B}(\Omega_1)$. *If* $\mu$ *strongly believes* $E \in \mathcal{B}(\Omega_1 \times \Omega_2)$ *and* $\text{proj}_{\Omega_1} E$ *is Borel, then* $\text{marg}_{\Omega_1} \mu$ *strongly believes* $\text{proj}_{\Omega_1} E$.

**Definition 4.5** *Say* $t_a \in T_a$ ***strongly believes*** $E_b \in \mathcal{B}(S_b \times T_b)$ *if* $\beta_a(t_a)$ *strongly believes* $E_b$.

Fix an event about Bob, viz. $E_b \in \mathcal{B}(S_b \times T_b)$. Write

$$\text{SB}_a(E_b) = S_a \times \{t_a \in T_a : t_a \text{ strongly believes } E_b\},$$

and $\text{CSB}_a(E_b) = E_b \cap \text{SB}_a(E_b)$. That is, $\text{SB}_a(E_b)$ is the event "Ann strongly believes $E_b$" and $\text{CSB}_a(E_b)$ is the event "Ann strongly believes $E_b$ and she is in fact correct." Given a product set $E \in \mathcal{B}(S_a \times T_a \times S_b \times T_b)$, viz. $E = E_a \times E_b$, write $\text{SB}(E) = \text{SB}_a(E_b) \times \text{SB}_b(E_a)$ and $\text{CSB}(E) = \text{CSB}_a(E_b) \times \text{CSB}_b(E_a)$.

Note, $\text{SB}(\cdot) = \text{SB}_a(\cdot) \times \text{SB}_b(\cdot)$ can be viewed as a mutual strong belief operator. Then, $\text{CSB}(\cdot) = \text{CSB}_a(\cdot) \times \text{CSB}_b(\cdot)$ is an auxiliary operator, which we will refer to as the "correct strong belief" operator. It will allow us to simplify the formulation of our epistemic assumptions. In particular, given a product set $E \in \mathcal{B}(S_a \times T_a \times S_b \times T_b)$, write $\text{CSB}^0(E) = E$ and, for each $m \geq 0$, define $\text{CSB}^{m+1}(E) = \text{CSB}(\text{CSB}^m(E))$. So,

$$\text{CSB}^1(E) = E \cap \text{SB}(E),$$

$$\text{CSB}^2(E) = \text{CSB}(E \cap \text{SB}(E)) = E \cap \text{SB}(E) \cap \text{SB}(E \cap \text{SB}(E)),$$

and so on. Note that

$$\mathrm{CSB}^{m+1}(E) = E \cap \bigcap_{n=0}^{m} \mathrm{SB}(\mathrm{CSB}^n(E)).$$

Now we can state the epistemic conditions of interest.

**Definition 4.6** *Say there is* ***rationality and common strong belief of rationality (RCSBR) at*** $(s_a, t_a, s_b, t_b)$ *if* $(s_a, t_a, s_b, t_b) \in \bigcap_m \mathrm{CSB}^m(R_a \times R_b)$.

# 5 Characterization Theorems

We now turn to characterizing RCSBR. For this it will be useful to introduce a **best reply correspondence**, viz. $\rho_a : \mathcal{C}(S_b) \to 2^{S_a}$, where $\rho_a(\mu_a)$ is the set of strategies that are sequentially optimal under $\mu_a$. We begin with extensive-form best response sets.

**Definition 5.1** *Call* $Q_a \times Q_b \subseteq S_a \times S_b$ *an* ***extensive-form best response set (EFBRS)*** *if, for each* $s_a \in Q_a$ *there is a CPS* $\mu_a \in \mathcal{C}(S_b)$ *so that:*

*(i)* $s_a \in \rho_a(\mu_a)$,

*(ii)* $\mu_a$ *strongly believes* $Q_b$, *and*

*(iii)* $\rho_a(\mu_a) \subseteq Q_a$.

*And similarly with a and b reversed.*

**Theorem 5.1** *Fix an extensive-form game* $\Gamma$.

*(i) For any* $\Gamma$*-based type structure,* $\mathrm{proj}_{S_a \times S_b} \bigcap_m \mathrm{CSB}^m(R_a \times R_b)$ *is an EFBRS.*

*(ii) Fix an EFBRS* $Q_a \times Q_b$. *There exists a* $\Gamma$*-based type structure, so that* $Q_a \times Q_b = \mathrm{proj}_{S_a \times S_b} \bigcap_m \mathrm{CSB}^m(R_a \times R_b)$.

To prove Theorem 5.1, it will be useful to point out a characterization of RCSBR: Let $R_a^0 = R_a$ (resp. $R_b^0 = R_b$). Inductively define $R_a^m$ (resp. $R_b^m$), so that $R_a^{m+1} = R_a^m \cap \mathrm{SB}_a(R_b^m)$ (resp. $R_b^{m+1} = R_b^m \cap \mathrm{SB}_b(R_a^m)$). Then, a standard induction argument gives that $\mathrm{CSB}^m(R_a \times R_b) = R_a^m \times R_b^m$, for each $m$. It follows that $\bigcap_m \mathrm{CSB}^m(R_a \times R_b) = \bigcap_m (R_a^m \times R_b^m)$. We make use of this below.

**Proof.** Begin by showing part (i) of the theorem. Fix a $\Gamma$-based type structure. If $\bigcap_m \mathrm{CSB}^m(R_a \times R_b) = \emptyset$ then the result is immediate. So, suppose $\bigcap_m \mathrm{CSB}^m(R_a \times R_b) \neq \emptyset$.

Fix $(s_a, s_b) \in \mathrm{proj}_{S_a \times S_b} \bigcap_m \mathrm{CSB}^m(R_a \times R_b)$. Then there exists $(t_a, t_b)$ such that

$$(s_a, t_a, s_b, t_b) \in \bigcap_m \mathrm{CSB}^m(R_a \times R_b) = \bigcap_m (R_a^m \times R_b^m).$$

We will show that the CPS $\mathrm{marg}_{S_b} \beta_a(t_a)$ satisfies conditions (i)-(iii) of an EFBRS, for the strategy $s_a$. A similar argument holds for $s_b$.

First note,

$$(s_a, t_a) \in \rho_a(\mathrm{marg}_{S_b} \beta_a(t_a)) \times \{t_a\} \subseteq R_a.$$

Now use the fact that $t_a$ strongly believes each $R_b^m$ to get that

$$\rho_a(\mathrm{marg}_{S_b} \beta_a(t_a)) \times \{t_a\} \subseteq \bigcap_m R_a^m.$$

So, $s_a \in \rho_a(\mathrm{marg}_{S_b} \beta_a(t_a)) \subseteq \mathrm{proj}_{S_a} \bigcap_m R_a^m$, establishing conditions (i) and (iii) of an EFBRS. Next note that, using the Conjunction Property of strong belief (Property 4.1), $\beta_a(t_a)$ strongly believes $\bigcap_m R_b^m$. Using the Marginalization Property (Property 4.2), $\mathrm{marg}_{S_a} \beta_a(t_a)$ strongly believes $\mathrm{proj}_{S_b} \bigcap_m R_b^m$. This establishes condition (ii) of an EFBRS.

Now turn to part (ii) of the Theorem. Fix an EFBRS $Q_a \times Q_b \neq \emptyset$. Let $T_a = Q_a$ and $T_b = Q_b$. Fix a type $t_a \in T_a = Q_a$. There is a CPS $\mu_a(t_a) \in \mathcal{C}(S_b)$ satisfying conditions (i)-(iii) of an EFBRS. Now construct a CPS $\beta_a(t_a) \in \mathcal{C}(S_b \times T_b, \mathcal{S}_b \otimes \mathcal{T}_b)$ as follows. If $Q_b \cap S_b(h) \neq \emptyset$, set $\beta_a(t_a)((t_b, t_b)|S_b(h) \times T_b) = \mu_a(t_a)(t_b|S_b(h))$ for each $t_b \in Q_b = T_b$. Next, fix some arbitrary element $t_b^* \in T_b$. If $Q_b \cap S_b(h) = \emptyset$, set $\beta_a(t_a)((s_b, t_b^*)|S_b(h) \times T_b) = \mu_a(t_a)(s_b|S_b(h))$ for each $s_b \in S_b$. (Note, $t_b^*$ is the same, for each information set with $Q_b \cap S_b(h) = \emptyset$.)

Indeed, each $\beta_a(t_a)$ is a CPS on $\mathcal{S}_b \otimes \mathcal{T}_b$. Note that conditions (i)-(ii) of a CPS are immediate. For condition (iii), fix an event $E_b$ and two information sets $h, i \in H_a$ with $E_b \subseteq S_b(h) \times T_b \subseteq S_b(i) \times T_b$. First, consider the case where $Q_b \cap S_b(h) \neq \emptyset$. In this case, $Q_b \cap S_b(i) \neq \emptyset$. So,

$$\begin{aligned}
\beta_a(t_a)(E_b|S_b(i) \times T_b) &= \mu_a(t_a)(\{t_b \in Q_b : (t_b, t_b) \in E_b\}|S_b(i)) \\
&= \mu_a(t_a)(\{t_b \in Q_b : (t_b, t_b) \in E_b\}|S_b(h)) \times \mu_a(t_a)(S_b(h)|S_b(i)) \\
&= \mu_a(t_a)(\{t_b \in Q_b : (t_b, t_b) \in E_b\}|S_b(h)) \times \mu_a(t_a)(Q_b \cap S_b(h)|S_b(i)) \\
&= \beta_a(t_a)(E_b|S_b(h) \times T_b) \times \beta_a(t_a)(S_b(h) \times T_b|S_b(i) \times T_b),
\end{aligned}$$

where the first and fourth lines follow from the construction, the second follows from the fact that $\mu_a(t_a)$ is a CPS, and the third line follows from the fact that $\mu_a(t_a)(Q_b|S_b(h)) = 1$ (since $Q_b \cap S_b(h) \neq \emptyset$ and $\mu_a(t_a)$ strongly believes $Q_b$). This establishes condition (iii) when $Q_b \cap S_b(h) \neq \emptyset$. So, suppose $Q_b \cap S_b(h) = \emptyset$ and recall $E_b \subseteq S_b(h) \times T_b$. If $Q_b \cap S_b(i) \neq \emptyset$, then $\mu_a(t_a)\left(\mathrm{proj}_{S_b} E_b|S_b(i)\right) = 0$ and $\mu_a(t_a)(S_b(h)|S_b(i)) = 0$. (This uses the fact that $\mu_a(t_a)(Q_b|S_b(i)) = 1$, which follows from strong belief.) So, here too,

$$\begin{aligned}
\beta_a(t_a)(E_b|S_b(i) \times T_b) &= \beta_a(t_a)(E_b|S_b(h) \times T_b) \times \beta_a(t_a)(S_b(h) \times T_b|S_b(i) \times T_b) \\
&= 0.
\end{aligned}$$

Finally, suppose $Q_b \cap S_b(i) = \emptyset$. Here,

$$
\begin{aligned}
\beta_a(t_a)(E_b|S_b(i)\times T_b) &= \mu_a(t_a)(\{s_b:(s_b,t_b^*)\in E_b\}|S_b(i)) \\
&= \mu_a(t_a)(\{s_b:(s_b,t_b^*)\in E_b\}|S_b(h))\times\mu_a(t_a)(S_b(h)|S_b(i)) \\
&= \beta_a(t_a)(E_b|S_b(h)\times T_b)\times\beta_a(t_a)(S_b(h)\times\{t_b^*\}|S_b(i)\times T_b) \\
&= \beta_a(t_a)(E_b|S_b(h)\times T_b)\times\beta_a(t_a)(S_b(h)\times T_b|S_b(i)\times T_b),
\end{aligned}
$$

as required.

We will conclude the proof by showing

$$Q_a = \bigcup_{t_a\in T_a}[\rho_a(\mathrm{marg}_{S_b}\beta_a(t_a))] \tag{5.1}$$

$$R_a^m = \bigcup_{t_a\in T_a}[\rho_a(\mathrm{marg}_{S_b}\beta_a(t_a))\times\{t_a\}] \quad \text{for each } m, \tag{5.2}$$

and likewise with $a$ and $b$ interchanged. Taken together, they give the desired result.

To show Equation 5.1: Recall, for each $t_a \in T_a = Q_a$, $\mu_a(t_a) = \mathrm{marg}_{S_b}\beta_a(t_a)$. So, it is immediate from the construction that $Q_a \subseteq \bigcup_{t_a\in T_a}\rho_a(\mathrm{marg}_{S_b}\beta_a(t_a))$. Conversely, fix any strategy $s_a$ in $\bigcup_{t_a\in T_a}\rho_a(\mathrm{marg}_{S_b}\beta_a(t_a))$. Then, there is a type $t_a \in T_a = Q_a$ so that $s_a$ is sequentially optimal under $\mu_a(t_a)(\cdot|\cdot)$. It follows from part (iii) of the definition of an EFBRS that $s_a \in Q_a$.

To show Equation 5.2: The proof is by induction on $m$. The Equation is immediate for $m = 0$. Assume the result holds for $m \geq 0$. In order to show that it holds for $m+1$, it suffices to show that each $t_a \in T_a$ strongly believes $R_b^m$. For this, fix an information set $h$ such that $R_b^m \cap [S_b(h)\times T_b] \neq \emptyset$. Observe that

$$
\begin{aligned}
[\mathrm{proj}_{S_b} R_b^m]\cap S_b(h) &= [\bigcup_{t_b\in T_b}\rho_b(\mathrm{marg}_{S_a}\beta_b(t_b))]\cap S_b(h) \\
&= Q_b \cap S_b(h).
\end{aligned}
$$

(The first equality follows from the induction hypothesis for $b$. The second equality follows from Equation 5.1.) Since $R_b^m \cap [S_b(h)\times T_b] \neq \emptyset$, it follows that $Q_b \cap S_b(h) \neq \emptyset$, and so $\mu_a(t_a)(Q_b|S_b(h)) = 1$. (Here, we use part (ii) of the definition of an EFBRS.) So, by construction, $\beta_a(t_a)(R_b^m|S_b(h)\times T_b) = 1$, as required. ■

Now, we turn to Directed Rationalizability. Let $\Delta_a$ (resp. $\Delta_b$) be a non-empty subset of $\mathcal{C}(S_b)$ (resp. $\mathcal{C}(S_a)$), i.e. a set of first-order beliefs of Ann (resp. Bob). Call $\Delta = \Delta_a \times \Delta_b$ a **set of first-order beliefs**. Set $S_a^{\Delta,0} = S_a$ and $S_b^{\Delta,0} = S_b$. Inductively define $S_a^{\Delta,m}$ and $S_b^{\Delta,m}$ as follows: Let $S_a^{\Delta,m+1}$ be the set of all $s_a \in S_a^{\Delta,m}$ so that there is some CPS $\mu_a \in \Delta_a$ with (i) $s_a \in \rho_a(\mu_a)$ and (ii) $\mu_a$ strongly believes $S_b^{\Delta,1},\ldots,S_b^{\Delta,m}$. And, likewise, with $a$ and $b$ interchanged.[7]

[7]This definition is as in [4, 1999]. It is a stronger requirement than the definition in Battigalli-Siniscalchi [9, 2003]. They put $s_a \in S_a^{\Delta,m+1}$ if $s_a \in S_a^{\Delta,m}$ and there is some CPS $\mu_a \in \Delta_a$ with (i) $s_a \in \rho_a(\mu_a)$ and (ii) $\mu_a$ strongly believes $S_b^{\Delta,m}$. Any set that satisfies the requirements here, also satisfies the requirements in [9, 2003]. But the

**Definition 5.2** *Call* $S_a^{\Delta} = \bigcap_{m \geq 0} S_a^{\Delta,m}$ *(resp.* $S_b^{\Delta} = \bigcap_{m \geq 0} S_b^{\Delta,m}$*) the* ***$\Delta$-rationalizable strategies of Ann*** *(resp.* ***Bob****). Call* $S_a^{\Delta} \times S_b^{\Delta}$ *the* ***$\Delta$-rationalizable strategy set.***

We use the phrase **Directed Rationalizability** to refer to the set of all $S_a^{\Delta} \times S_b^{\Delta}$. So, for a given game $\Gamma$, the Directed Rationalizability concept gives

$$\{S_a^{\Delta} \times S_b^{\Delta} : \Delta = \Delta_a \times \Delta_b \subseteq \mathcal{C}(S_b) \times \mathcal{C}(S_a)\}.$$

Note, since the sets $S_a^{\Delta,m} \times S_b^{\Delta,m}$ form a decreasing sequence and $S_a \times S_b$ is finite, there is some (finite) $M$ so that $S_a^{\Delta} \times S_b^{\Delta} = S_a^{\Delta,M} \times S_b^{\Delta,M}$. Also, note that, for a given set of first-order beliefs, viz. $\Delta = \Delta_a \times \Delta_b$, the $\Delta$-rationalizable strategy set may be empty.

Beginning from the lady's choice example, we can use the type structure to construct an associated set of first-order beliefs $\Delta$ and this set of first-order beliefs $\Delta$ can be used to perform $\Delta$-rationalizability. The output is the EFBRS we identified earlier. But, the lady's choice convention had a particular feature: it was a restriction on first-order beliefs and a requirement that the restriction be "transparent" to the players. So, the only restriction on second-order beliefs (i.e., beliefs about strategy the other player chooses and the other player's the first-order beliefs) was the requirement that, at each information set, Ann must believe that Bob believes she will play *Up.* And so on. It was this transparency of (only) first-order restrictions that allowed us to directly compute the associated Directed Rationalizability set.

More generally, when we begin from a given type structure, we impose substantive assumptions about which beliefs players do versus do not consider possible. These assumptions may correspond to restrictions (only) on players' first-order beliefs which are transparent to the players. But, they need not—they may involve additional restrictions on higher-order beliefs. And, if they do, the procedure outlined above fails.

To see the failure, begin with an epistemic type structure and use the structure itself to form the set $\bar{\Delta} = \bar{\Delta}_a \times \bar{\Delta}_b$. Specifically, for each type $t_a \in T_a$, consider the marginal of $\beta_a(t_a)$ on $S_b$. These CPS's form the set $\bar{\Delta}_a$. Construct the set $\bar{\Delta}_b$ analogously. Here, the strategies that survive one round of $\bar{\Delta}$-rationalizability are exactly the strategies that are consistent with $\mathrm{R0SBR}_a \times \mathrm{R0SBR}_b$. But, on round two, we lose the equivalence: If $\beta_a(t_a)$ strongly believes the event "Bob is rational," then the marginal of $\beta_a(t_a)$ will also strongly believe that "Bob chooses a strategy consistent with one round of elimination of $\bar{\Delta}$-rationalizability." (Here, we use a marginalization property of strong belief, plus the round-one equivalence.) But, the converse need not hold. So, the strategies that survive two rounds of $\bar{\Delta}$-rationalizability may strictly contain the R1SBR strategies. And, on round three, we loose the inclusion. If the CPS $\beta_a(t_a)$ strongly believes the R1SBR event for Bob, then the marginal of $\beta_a(t_a)$ will also strongly believe that "Bob chooses a strategy consistent with R1SBR."

converse does not hold. (See Battigalli-Prestipino [5, 2010] for an example.) Thus, using Theorem 5.1 here, it can be shown that the definition of [9, 2003] is conceptually incorrect. (Battigalli-Prestipino [5, 2010] point out that the two definitions are equivalent when $\Delta$ satisfies a "closedness under composition" condition. Since [9, 2003] focus on the case where this condition is satisfied, their results hold with the definition given here.)

But, recall, the strategies consistent with R1SBR may be strictly contained in the strategies that survive two rounds of $\bar{\Delta}$-rationalizability. So there may be information sets consistent with this latter event, but not the former. This implies that, even if $\beta_a(t_a)$ strongly believes the R1SBR event for Bob, it need not strongly believe that Ann's behavior is consistent with two rounds of $\bar{\Delta}$-rationalizability. (This is an instance of the fact that strong belief is not monotonic.) As such, we loose any relationship between the RCSBR strategies and the $\bar{\Delta}$-rationalizable strategy set. In fact, Appendix C illustrates an example where the RCSBR strategy set and the $\bar{\Delta}$-rationalizable strategy set are disjoint.

But, there is another route, that instead uses the EFBRS properties to form a set $\Delta = \Delta_a \times \Delta_b$ of first-order beliefs. Fix an epistemic structure. The RCSBR strategies form an EFBRS, viz. $Q_a \times Q_b$. For each $s_a \in Q_a$, we have some CPS $\mu_a(s_a)$ satisfying the conditions of an EFBRS. Take $\Delta_a$ to be the set of such CPS's, i.e., one for each $s_a \in Q_a$, and construct $\Delta_b$ similarly. Now we do have an equivalence between the RCSBR strategies and the $\Delta$-rationalizable strategies. More precisely, for each $m \geq 1$, $Q_a \times Q_b$ is the set of strategies that survives $m$-rounds of elimination of $\Delta$-rationalizability. The case of $m = 1$ follows from properties (i) and (iii) of an EFBRS. The case of $m = 2$ uses condition (ii) of an EFBRS. And so on, by induction.

**Proposition 5.1** *Fix an extensive-form game* $\Gamma$.

*(i) Given an EFBRS, viz.* $Q_a \times Q_b$, *there exists a set of first-order beliefs, viz.* $\Delta = \Delta_a \times \Delta_b$, *so that* $S_a^\Delta \times S_b^\Delta = Q_a \times Q_b$.

*(ii) Given a set of first-order beliefs, viz.* $\Delta = \Delta_a \times \Delta_b$, $S_a^\Delta \times S_b^\Delta$ *is an EFBRS.*

Thus, in conjunction with Theorem 5.1, we have the following Characterization Theorem.

**Corollary 5.1** *Fix an extensive-form game* $\Gamma$.

*(i) For any* $\Gamma$*-based type structure, there exists a set of first-order beliefs, viz.* $\Delta = \Delta_a \times \Delta_b$, *so that* $S_a^\Delta \times S_b^\Delta = \mathrm{proj}_{S_a \times S_b} \bigcap_m \mathrm{CSB}^m (R_a \times R_b)$.

*(ii) Fix a set of first-order beliefs, viz.* $\Delta_a \times \Delta_b$. *Then there exists a* $\Gamma$*-based structure, so that* $S_a^\Delta \times S_b^\Delta = \mathrm{proj}_{S_a \times S_b} \bigcap_m \mathrm{CSB}^m (R_a \times R_b)$.

Now for the proof.

**Proof of Proposition 5.1.** Begin with part (i). Fix $Q_a \times Q_b$. For each $s_a \in Q_a$, there exists a corresponding CPS $\mu_a(s_a) \in \mathcal{C}(S_b)$ satisfying conditions (i)-(iii) of an EFBRS for $Q_a \times Q_b$. Take $\Delta_a$ so that, for each $s_a \in Q_a$, $\Delta_a$ contains exactly one such CPS $\mu_a(s_a)$. There are no other CPS's in $\Delta_a$. Define $\Delta_b$ analogously. We will show that, for each $m \geq 1$, $S_a^{\Delta,m} \times S_b^{\Delta,m} = Q_a \times Q_b$. This will establish the result.

The proof is by induction. Begin with $m = 1$. Certainly $Q_a \subseteq S_a^{\Delta,1}$. Fix $s_a \in S_a^{\Delta,1}$. Then there exists some $\mu_a \in \Delta_a$ so that $s_a$ is sequentially optimal under $\mu_a$. This CPS $\mu_a$ is associated with some $r_a \in Q_a$, i.e., so that $r_a$ and $\mu_a$ jointly satisfy conditions (i)-(iii) of an EFBRS. Now apply condition (iii) of an EFBRS to get that $s_a \in Q_a$. And, likewise, for $b$.

Now assume $S_a^{\Delta,m} \times S_b^{\Delta,m} = Q_a \times Q_b$ for $m \geq 2$. We will show it also holds for $m+1$. Fix $s_a \in Q_a = S_a^{\Delta,m}$. Then, using the construction of $\Delta_a$, there exists some $\mu_a \in \Delta_a$ satisfying conditions (i)-(ii) of an EFBRS for $Q_a \times Q_b$, so that $s_a \in \rho_a(\mu_a)$ and $\mu_a$ strongly believes $Q_b = S_b^{\Delta,m}$. So, certainly, $Q_a \subseteq S_a^{\Delta,m+1}$. Conversely, fix some $s_a \in S_a^{\Delta,m+1}$. Then, there exists a CPS $\mu_a \in \Delta_a$ so that $s_a \in \rho_a(\mu_a)$ and $\mu_a$ strongly believes $S_b^{\Delta,m}$. Again, since each element of $\Delta_a$ satisfies conditions (i)-(iii) of an EFBRS for some $r_a \in Q_a$, it follows that $\rho_a(\mu_a) \subseteq Q_a$, and so $s_a \in Q_a$.

Now turn to part (ii). Fix some set of first-order beliefs, viz. $\Delta = \Delta_a \times \Delta_b$. Note, there exists some $M$ with $S_a^{\Delta} \times S_b^{\Delta} = S_a^{\Delta,M} \times S_b^{\Delta,M} = S_a^{\Delta,M+1} \times S_b^{\Delta,M+1}$. Fix $s_a \in S_a^{\Delta} = S_a^{\Delta,M+1}$. We can find a CPS $\mu_a \in \Delta_a$ so that $s_a \in \rho_a(\mu_a)$ and $\mu_a$ strongly believes each $S_b^{\Delta,m}$ for $m \leq M$. In particular, $\mu_a$ strongly believes $S_b^{\Delta,M} = S_b^{\Delta}$. Thus, $s_a$ satisfies conditions (i)-(ii) of an EFBRS for $Q_a \times Q_b = S_a^{\Delta} \times S_b^{\Delta}$. Moreover, if $r_a \in \rho_a(\mu_a)$, then $r_a$ is optimal under a CPS that strongly believes each $S_b^{\Delta,m}$, for $m \leq M$. As such, $r_a \in S_a^{\Delta,M+1} = S_a^{\Delta}$. Therefore also condition (iii) of an EFBRS is satisfied. A similar argument applies to $b$. Therefore $S_a^{\Delta} \times S_b^{\Delta}$ is an EFBRS. ■

Let us comment on the proof. Begin with some finite set of first-order beliefs, viz. $\Delta = \Delta_a \times \Delta_b$. Proposition 5.1(ii) says that $S_a^{\Delta} \times S_b^{\Delta}$ is an EFBRS. Conversely, begin with some EFBRS. The proof of Proposition 5.1(i) says that we can find a finite set of first-order beliefs, viz. $\Delta = \Delta_a \times \Delta_b$, so that $S_a^{\Delta} \times S_b^{\Delta}$ is this EFBRS. With this in mind:

**Remark 5.1** *Fix a game tree $\Gamma$. The Directed Rationalizability set is*

$$\{S_a^{\Delta} \times S_b^{\Delta} : \Delta = \Delta_a \times \Delta_b \subseteq \mathcal{C}(S_b) \times \mathcal{C}(S_a)\} = \{S_a^{\Delta} \times S_b^{\Delta} : \Delta = \Delta_a \times \Delta_b \text{ is finite}\}.$$

Thus, using the EFBRS properties, we can see that we only need to compute the $\Delta$-rationalizable sets for finite sets of first-order beliefs.

# 6 Analyzing Games

We now turn to the analysis of games. In Section 1.6, we used the EFBRS concept to analyze Battle of the Sexes with an Outside Option. Now, we turn to analyze other games of interest. This will help us, more generally, better understand what the EFBRS does vs. does not give in games of interest. The approach will be to make use of Properties (i)-(iii) (usually only Properties (i)-(ii)) of the EFBRS definition, and not the equivalent Directed Rationalizability definition.

Let us begin with the Centipede game .

**Example 6.1** *Consider the three-legged Centipede game, given in Figure 6.1 below.*

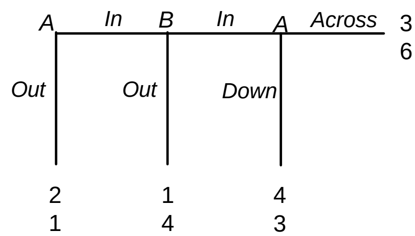


*Figure 6.1*

*Here, the EFBRS's are* $\{Out\} \times \{Out\}$ *and* $\{Out\} \times \{Out, In\}$.

*Notice, we cannot have an EFBRS where Ann plays In at the first node. To see this, suppose otherwise, i.e., there exists an EFBRS* $Q_a \times Q_b$ *and a strategy* $s_a \in Q_a$ *where* $s_a$ *plays In at the first node. Note, by condition (i) of an EFBRS, we must have that* $Q_a \subseteq \{Out, In\text{-}Down\}$, *so that* $s_a = In\text{-}Down$. *Now, fix* $s_b \in Q_b$ *and note that* $s_b$ *is sequentially optimal under a CPS that strongly believes* $Q_a$. *Then, at Bob's information set, this CPS must assign probability one to In-Down. Since* $s_b$ *is sequentially optimal under this CPS,* $s_b = Out$. *So, we have that* $Q_b = \{Out\}$. *But, then, In-Down cannot simultaneously satisfy conditions (i)-(ii) of an EFBRS.*

The argument we have presented for the three-legged Centipede is more general. In particular, fix an EFBRS for an $n$-legged Centipede game. Under the EFBRS, the first player chooses *Out*. This will be a consequence of Proposition 7.1(i) to come.

**Example 6.2** *Figure 6.2 gives the Prisoner's Dilemma. Consider the* 3*-repeated version of the game.*

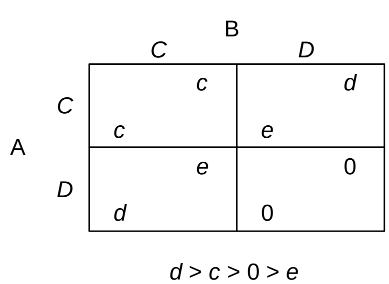


*Figure 6.2*

*Let* $Q_a \times Q_b$ *be a nonempty EFBRS. Then each* $(s_a, s_b) \in Q_a \times Q_b$ *results in the Defect-Defect* *path.*[8]

*Let us give an intuition: First, note, each strategy $s_a \in Q_a$ (resp. $s_b \in Q_b$) is sequentially justifiable. (This is condition (i).) As such, $s_a$ (resp. $s_b$) plays Defect in the last period, at each history allowed by $s_a$ (resp. $s_b$). Now, consider a second period information set $h$, where $s_a \in S_a(h)$ and $Q_b \cap S_b(h) \neq \emptyset$. By conditions (i)-(ii) of an EFBRS, $s_a$ must be sequentially optimal under a CPS $\mu_a(s_a)$ with $\mu_a(s_a)(Q_b|S_b(h)) = 1$. Note, then, conditional upon $h$, $\mu_a(s_a)$ assigns probability one to Bob defecting in the third period, irrespective of Ann's play. As such, $s_a$ plays $D$ at $h$. And, likewise, with $a$ and $b$ reversed.*

*Now turn to the first period, and suppose, contra hypothesis, there is some $s_a \in Q_a$ so that $s_a$ chooses $C$ as $\phi$. Note, for each $s_a \in Q_b$, $(s_a, s_b)$ results in the Defect-Defect path in periods two and three. So, Ann's expected payoffs from $s_a$ corresponds to her first period expected payoffs from playing $s_a$. Now note that, the Defect-always strategy yields a strictly higher expected payoff in the first period and an expected payoff of at least zero in subsequent periods. As such, this contradicts $s_a$ being optimal under $\mu_a(s_a)(\cdot|S_b)$.*

An analogous result holds for the $N$-repeated Prisoner's Dilemma, for $N$ finite. The proof is given in Appendix D.

Let us pause to take stock of the answers we have seen. First, in Battle of the Sexes with the Outside Option, we got that either (i) Bob plays *Out* or (ii) Bob plays *In-Right* and Ann plays *Down*. Each of these were subgame perfect paths of play. In Centipede, we saw that we get the backward induction path (but not necessarily the backward induction strategies). And, likewise, in the Finitely Repeated Prisoner's Dilemma, we said that we get the unique Nash (and so subgame perfect) path, namely where each player *Defects* in all periods.

Note, then, in each of these cases, the outcomes allowed by an EFBRS coincide with the outcomes allowed by some subgame perfect equilibrium (SPE). This raises the question: what is the relationship between the EFBRS concept and the SPE concept? Are the two concepts equivalent? If so, then we have a good idea what the EFBRS concept delivers (in games of interest), since we have a good idea about what SPE delivers.

We will see that, in a particular class of games, any pure-strategy SPE corresponds to some EFBRS. Each of the examples we mentioned is contained in this class of games. But the EFBRS and SPE concepts do not coincide. Let us begin with the positive result and then turn to the negative results. First, we need a definition.

**Definition 6.1** *Say a game $\Gamma$ has **observable actions** if each information set is a singleton.*

To understand the definition, recall, in our set-up, both $a$ and $b$ have a choice at each history. (Of course, it may be the case that only one of the players is active.) So, a game with observable

[8] In the once or twice repeated Prisoner's Dilemma we have a stronger result: If $(s_a, s_b)$ is contained in an EFBRS, then each of $s_a$ and $s_b$ specify *Defect* at each history.

actions is one where the players begin by making simultaneous choices, learn the realization of the choices and then perhaps make simultaneous choices, etc., until a terminal history is reached.

Given distinct terminal histories, viz. $z$ and $z'$, we can write $z = \left(x, c^1, \ldots, c^K\right)$ and $z' = \left(x, d^1, \ldots, d^L\right)$, where $x$ is the last common predecessor of $z$ and $z'$, i.e., $c^1 \neq d^1$. (Recall, $c^k = \left(c^k_a, c^k_b\right)$ and $d^l = \left(d^l_a, d^l_b\right)$.) Now:

**Definition 6.2** *Fix a game with observable actions and two distinct terminal nodes* $z = \left(x, c^1, \ldots, c^K\right)$ *and* $z' = \left(x, d^1, \ldots, d^L\right)$*. Say* **a *is decisive for*** $(\mathbf{z}, \mathbf{z}')$ *if* $a$ *moves at* $x$*,* $c^1_a \neq d^1_a$*, and* $c^1_b = d^1_b$*. And, likewise, with* $a$ *and* $b$ *interchanged.*

**Definition 6.3 (Battigalli [3, 1997])** *A game satisfies* ***no relevant ties (NRT)*** *if whenever* $a$ *(resp.* $b$*) is decisive for* $(z, z')$*,* $\Pi_a(z) \neq \Pi_a(z')$*.*

A game with no ties satisfies NRT, but the converse does not hold. Reny's [25, 1993; Figure 1] Take-It-Or-Leave-It game is one such example.

Fix a strategy $s_a$ and write $[s_a]$ for the set of all $r_a$ that induce the same plan of action as $s_a$, i.e., the set of all $r_a$ so that $\zeta(r_a, \cdot) = \zeta(s_a, \cdot)$. And, likewise, define $[s_b]$.

**Proposition 6.1** *Fix a game* $\Gamma$ *with observable actions and a pure-strategy SPE, viz.* $(s_a, s_b)$*.*

*(i) There is an EFBRS, viz.* $Q_a \times Q_b$*, so that* $[s_a] \times [s_b] \subseteq Q_a \times Q_b$*.*

*(ii) If* $\Gamma$ *satisfies NRT, then* $[s_a] \times [s_b]$ *is an EFBRS.*

Appendix D proves a result somewhat more general than Proposition 6.1. Note, each of the examples we have seen satisfied both observable actions and NRT. In each of these examples, any pure-strategy subgame perfect equilibrium $(s_a, s_b)$ belongs to an EFBRS, where the EFBRS only allows the terminal node $\zeta(s_a, s_b)$. This fits with part (ii) of the Proposition. Part (i) says that, even if the game fails NRT, $(s_a, s_b)$ will still be contained in some EFBRS—but now the EFBRS may also allow new paths. Let's take a simple example of this.

**Example 6.3** *The game in Figure 6.3. Here,* $(In, Across)$ *is a pure-strategy SPE. There is an EFBRS, viz.* $Q_a \times Q_b$*, with* $\{In\} \times \{Across\} \subseteq Q_a \times Q_b$*, e.g.,* $\{In\} \times \{Across, Down\}$*. (Of course, part (i) of Proposition 6.1 says there must be some such EFBRS.) But every EFBRS, viz.* $Q_a \times Q_b$*, must have* $Q_b = \{Across, Down\}$*. (Here we use condition (iii) of an EFBRS.) So,* $\{In\} \times \{Across\}$

*is not an EFBRS.*

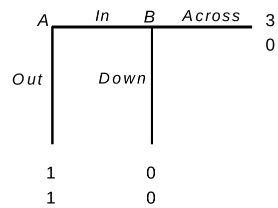


*Figure 6.3*

Proposition 6.1 tells us that, in certain games, a pure-strategy subgame perfect equilibrium can be viewed as an EFBRS. This is one step toward analyzing games under the EFBRS concept. But, it is not the end of the matter. In general, the two concepts are not equivalent. For a given game, there may be a pure-strategy subgame perfect equilibrium whose outcome is precluded by any EFBRS. (Of course, per Proposition 6.1, this can only occur in games that do not have observable actions.) And, conversely, a given EFBRS may allow outcomes which are precluded by any subgame perfect equilibrium—even any subgame perfect equilibrium in behavioral strategies. (We will see that this can happen in a game with observable actions and NRT.) The next examples demonstrate these points.

**Example 6.4** *Consider the game in Figure 6.4, which fails the observable actions condition. It is obtained from the game in Figure 1.4 by two transformations. First, the simultaneous move subgame is transformed into one where Ann moves first and then Bob moves not knowing Ann's choice. Second, two of Ann's decision nodes are coalesced.*

*Here,* $(Out, Right)$ *is a pure strategy subgame perfect equilibrium. But, again applying the argument in Section 1.5, we get that Out is not contained in any EFBRS.*[9]

[9] Note that, unlike the subgame perfect concept, the EFBRS concept is invariant to coalescing decision nodes.

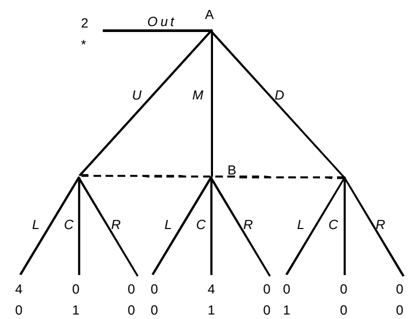


Figure 6.4

**Example 6.5** *Consider the coordination game in Figure 6.5.*

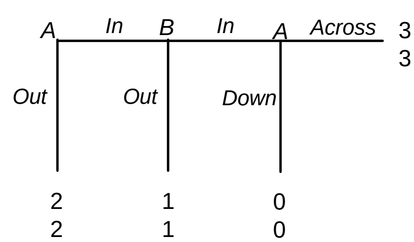


*Figure 6.5*

*Here, there are three EFBRS's, namely* $\{Out\} \times \{Out\}$, $\{Out\} \times \{Out, In\}$, *and* $\{In\text{-}Across\} \times \{In\}$. *The unique subgame perfect equilibrium is* $(In\text{-}Across, In)$, *which results in the* $(3,3)$ *outcome. Indeed, this profile induces an EFBRS. But, there are two EFBRS's which involve Ann playing Out at the initial node. This results in an outcome, viz.* $(2,2)$, *precluded by any subgame perfect equilibrium (even in behavioral strategies).*

Taken together with the Main Theorem (Theorem 5.1), Example 6.5 says that a non-backward induction outcome, namely $(2,2)$, is consistent with RCSBR. To understand this better, note that *Out* is the unique best response for Ann, under a CPS that assigns probability one to *Out* at the initial node. So, if each type of Ann assigns probability one to $\{Out\} \times T_b$, then conditional upon Bob's node begin reached, he must conclude that Ann is irrational. In this case, Bob may very well "think" that Ann is playing *In-Down*. If a type $t_b$ of Bob does maintain such a hypothesis, *Out* is a unique best response for $t_b$.

# 7 Perfect Information Games

In light of Example 6.5, even in games with observable actions (and NRT), the EFBRS and SPE concepts do not coincide. Thus, even in these games, we cannot use the SPE concept to analyze the consequences of context-dependent forward induction reasoning.

Now, we turn to a particular class of games with observable actions—namely, perfect information games (i.e., games with observable actions and with at most one active player at each information set). We've seen some examples of perfect-information games. Let's focus on Examples 6.1 and 6.5. In the former case, each EFBRS yields the backward induction path (and so the backward induction outcome). Of course, for that game, the Nash and backward induction paths coincide. On the other hand, in Example 6.5, one EFBRS corresponds to backward induction, but others do not. However, there we do get that the EFBRS paths correspond (exactly) to the Nash paths (and so Nash outcomes) of the game.

The examples suggest there may be a connection between EFBRS's and Nash outcomes, at least for perfect-information (PI) games. (Of course, for non-PI games, an EFBRS may give non-Nash outcomes.) Indeed, there will be a connection, for PI games satisfying a "no ties" condition.

**Definition 7.1 (Brandenburger-Friedenberg [15, 2004])** *A game satisfies the* ***single payoff condition (SPC)*** *if whenever* $a$ *(resp.* $b$*) is decisive for* $(z, z')$ *and* $\Pi_a(z) = \Pi_a(z')$*, then* $\Pi_b(z) = \Pi_b(z')$*.*

Of course, a game satisfying NRT also satisfies SPC. Yet, many games of interest satisfy SPC, but fail NRT, e.g., zero sum games. In perfect-information games, SPC is equivalent to "transference of decision-maker indifference" (Marx-Swinkels [21, 1997]).[10]

Now let us state the connection:

**Proposition 7.1**

*(i) Fix a PI game* $\Gamma$ *satisfying SPC. If* $Q_a \times Q_b$ *is an EFBRS then, there exists a pure-strategy Nash equilibrium, viz.* $(s_a, s_b)$*, so that each profile in* $Q_a \times Q_b$ *is outcome equivalent to* $(s_a, s_b)$*.*

*(ii) Fix a PI game* $\Gamma$ *satisfying NRT. If* $(s_a, s_b)$ *is a pure-strategy Nash equilibrium in sequentially justifiable strategies, then there is an EFBRS, viz.* $Q_a \times Q_b$*, so that* $(s_a, s_b) \in Q_a \times Q_b$*.*

The proof can be found in Appendix E. Taken together Theorem 5.1 and Proposition 7.1 give:

**Corollary 7.1**

*(i) Fix a PI game* $\Gamma$ *satisfying SPC, and an epistemic type structure. If there is RCSBR at the state* $(s_a, t_a, s_b, t_b)$*, then* $(s_a, s_b)$ *is outcome equivalent to a pure-strategy Nash equilibrium.*

[10] The SPC is a condition stated on the tree. Transference of decision-maker indifference is stated on the matrix. Here, it will be convenient to use a condition defined on the tree.

*(ii) Fix a PI game $\Gamma$ satisfying NRT, and a pure-strategy Nash equilibrium, viz. $(s_a, s_b)$, in sequentially justifiable strategies. Then, there exists an epistemic structure and a state thereof, viz. $(r_a, t_a, r_b, t_b)$, at which there is RCSBR and $(r_a, r_b) = (s_a, s_b)$.*

Why the connection between EFBRS's and Nash equilibria? Recall, the Preliminary Observation in Aumann-Brandenburger [1, 1995]: If each player is "rational," i.e., plays a best response, and places probability one on the actual strategy choices by the other player, then the strategy choices constitute a Nash equilibrium. In a PI game satisfying SPC, RCSBR imposes a form of correct beliefs about the actual outcomes that will obtain. Let us recast this at the level of the solution concept: In a PI game satisfying SPC, each strategy profile in a given EFBRS is outcome equivalent. (This will be Lemma E2 in Appendix E.) So, along the path of play, the associated CPS('s) must assign probability one to a particular outcome—the outcome associated with the EFBRS, i.e., the "correct" outcome. (This uses condition (ii) of an EFBRS.) With this, we get a Nash outcome (but not necessarily the Nash strategies).

This was the intuition for part (i) of Corollary 7.1. The proof follows the proof of Proposition 6.1a in Brandenburger-Friedenberg [15, 2004], though now making use of the EFBRS properties. (The proof in [15, 2004] makes use of properties of self-admissible sets. See 8c below.) Indeed, we only use properties (i)-(ii) of Definition 5.1.

The converse, i.e., part (ii), is novel. (In particular, both the 'no ties' condition and the proof are quite different from the analysis in [15, 2004].) A Nash equilibrium in sequentially justifiable strategies will, in general, satisfy conditions (i)-(ii) of an EFBRS. However, it may fail the maximality criterion. Indeed, the proof makes use of all three properties of Definition 5.1. See Appendix E.

The no ties conditions are important for both directions of Proposition 7.1. We explain why, by way of a number of examples.

**Example 7.1** *Return to Example 6.3, which fails SPC. There, we saw that $(In, Down)$ is contained in an EFBRS. But, it is not outcome equivalent to a pure-strategy Nash equilibrium.*

To better understand this last example, notice: When Bob moves, he is indifferent between $In$ and $Out$. Now turn to a type of Ann that strongly believes Bob is rational. This type has a correct belief about what Bob's payoffs will be if she plays $In$. But, because the game fails SPC, she may have an incorrect belief about what her own payoff will be if she plays $In$. As such, a Nash outcome need not obtain.

The next example shows that part (ii) of Proposition 7.1 may be false, if we replace the NRT condition with the SPC condition.

**Example 7.2** *Consider the game in Figure 7.1, which satisfies SPC.*

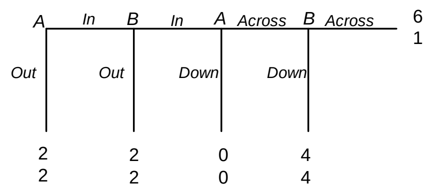


*Figure 7.1*

*Here, $(Out, Out)$ is a Nash equilibrium in sequentially justifiable strategies. But, if $Q_a \times Q_b$ is a (nonempty) EFBRS, then $Q_a \times Q_b = \{In\text{-}Across\} \times \{In\text{-}Down\}$. To see this, let $Q_a \times Q_b \neq \emptyset$ be an EFBRS and note that $Q_a \subseteq \{Out, In\text{-}Across\}$ and $Q_b \subseteq \{Out, In\text{-}Down\}$. (The strategy In-Down for Ann is dominated at her second information set, and the strategy In-Across for Bob is dominated at his second information set.) Note, too, that In-Across is a weakly dominant strategy for Ann. So, condition (iii) of an EFBRS implies that $In\text{-}Across \in Q_a$. It follows that, if $\mu_b$ strongly believes $Q_a$, then $\mu_b$ must assign probability one to In-Across conditional on the event "Ann plays In." So, In-Down is Bob's only sequential best response to any CPS that strongly believes $Q_a$. This implies that $Q_b = \{In\text{-}Down\}$, and so $Q_a = \{In\text{-}Across\}$.*

Do note: In the above example, $\{(Out, Out)\}$ is disjoint from any EFBRS. While it satisfies conditions (i)-(ii) of an EFBRS, it fails condition (iii): If $(Out, Out)$ is played, Ann gets a payoff of 2. But, by going $In$, she can also assure herself an expected payoff of at least 2. As such, condition (iii) requires that we include *In-Across*.

To better understand what is going on, let us recast this at the epistemic level: If $(Out, t_a)$ is rational, so is $(In\text{-}Across, t_a)$. With this, if Bob strongly believes that Ann is rational, then, when his first information set is reached, he must maintain a hypothesis that Ann is playing *In-Across*—that is, he must maintain a hypothesis that Ann is playing a particular strategy that is not in $Q_a = \{Out\}$. As such, $Out$ cannot be a best response for Bob.

The key is that the rationality of $(Out, t_a)$ has implications for Ann's rationality at information sets precluded by $Out$. Notice, this happens because Ann is indifferent between the terminal nodes reached by $(Out, Out)$ and $(In\text{-}Across, Out)$. (If Ann's payoffs from $(In\text{-}Across, Out)$ were strictly less than 2, $(Out, t_a)$ can be rational without $(In\text{-}Across, t_a)$ being rational. Similarly, if Ann's payoffs from $(In\text{-}Across, Out)$ were strictly greater than 2, then $(Out, Out)$ would not be a Nash Equilibrium.) This is where the NRT condition comes in—it says that, if Ann is decisive between two terminal nodes (as she is here), then she cannot be indifferent between those nodes.

Finally, notice there is a gap between parts (i)-(ii) of Proposition 7.1. In particular, part (i) says that starting from an EFBRS we can get a pure Nash outcome, while part (ii) says that starting from a sequentially justifiable pure-strategy Nash equilibrium, we can get an EFBRS.

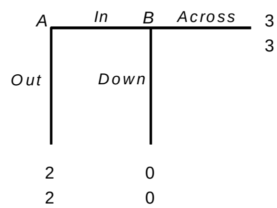


Figure 7.2

We cannot improve part (ii) to say that, starting from any pure Nash equilibrium, we get an EFBRS. To see this, refer to Figure 7.2. There is a unique EFBRS, namely $\{In\} \times \{Across\}$. That said, the pair $(Out, Down)$ is a Nash equilibrium—of course, it is not a Nash equilibrium in sequentially justifiable strategies.

Can we improve part (i) to say that, starting from an EFBRS, we get a pure-strategy Nash equilibrium in sequentially justifiable strategies? We do not know. In Appendix E, we elaborate on the issue. However, we note that, starting from an EFBRS, we can get a mixed-strategy Nash equilibrium that satisfies a "sequential justifiability" condition. (We'll make the condition precise below.)

Consider a pure strategy profile $(s_a, s_b)$ and a mixed strategy profile $(\varpi_a, \varpi_b) \in \mathcal{P}(S_a) \times \mathcal{P}(S_b)$. Call $(s_a, s_b)$ and $(\varpi_a, \varpi_b)$ **outcome equivalent** if $\pi(s_a, s_b) = \pi(\varpi_a, \varpi_b)$. Likewise, call $Q_a \times Q_b \subseteq S_a \times S_b$ and $(\varpi_a, \varpi_b) \in \mathcal{P}(S_a) \times \mathcal{P}(S_b)$ **outcome equivalent** if each $(s_a, s_b) \in Q_a \times Q_b$ is outcome equivalent to $(\varpi_a, \varpi_b)$. Then:

**Proposition 7.2** *Fix a PI game satisfying SPC. If $Q_a \times Q_b$ is an EFBRS, then there exists a mixed-strategy Nash equilibrium, viz. $(\sigma_a, \sigma_b)$, so that:*

*(i) $Q_a \times Q_b$ is outcome equivalent to $(\sigma_a, \sigma_b)$, and*

*(ii) each $s_a \in \operatorname{Supp} \sigma_a$ (resp. $s_b \in \operatorname{Supp} \sigma_b$) is sequentially justifiable.*

Proposition 7.2 gives that, if we begin with an EFBRS, we can construct an associated mixed-strategy Nash equilibrium. The Nash equilibrium has the property that each strategy in its support is sequentially justifiable. But, it is important to note that this does not necessarily give that the mixed-strategy itself is sequentially justifiable.[11] More to the point: Given a PI game satisfying SPC and some mixed-strategy Nash equilibrium, viz. $(\sigma_a, \sigma_b)$, does there exist some pure-strategy

[11] In non-PI games, we can construct a mixed-strategy Nash equilibrium, viz. $(\sigma_a, \sigma_b)$, where each strategy in the support of $\sigma_a$ and $\sigma_b$ is sequentially justifiable, but $\sigma_a$ is itself not sequentially justifiable. The question remains whether or not the same can occur in PI games.

Nash equilibrium, viz. $(s_a, s_b)$, so that $s_a$ (resp. $s_b$) is contained in the support of $\sigma_a$ (resp. $\sigma_b$)? If so, using Proposition 7.2, we get that starting from an EFBRS, there is a pure-strategy Nash equilibrium in sequentially justifiable strategies. But, this too is not known.

# 8 Discussion Section

In this section, we discuss some conceptual aspects of the paper, as well as some extensions.

**a. The Question:** Here, we study context-dependent forward induction reasoning. We focus on the case where the analyst does not know the specific context within which the game is played. With this in mind, we ask: Can we characterize RCSBR (i.e., across all type structures)? Indeed we can. We have seen that the EFBRS concept does just that. Or, alternatively, that the Directed Rationalizability solution concept characterizes RCSBR across all type structures.

Note, carefully, that Battigalli [4, 1999] and Battigalli-Siniscalchi [9, 2003] introduced Directed Rationalizability as an answer to a different question: They were interested in the case where the analyst knows the particular context, and the context only imposes a "transparent restriction" on players' first-order beliefs.[12] Specifically, the analyst is given a set of first-order beliefs, viz. $\Delta_a \times \Delta_b$, which satisfies two conditions: (i) each type $t_a$ has $\text{marg}_{S_b} \beta_a(t_a)$ contained in $\Delta_a$ and (ii) for each CPS $\mu_a$ on $S_b \times T_b$ with $\text{marg}_{S_b} \mu_a \in \Delta_a$, there is a type $t_a$ with $\beta_a(t_a) = \mu_a$. And, likewise, with $a$ and $b$ interchanged. (Condition (ii) ensures that the type structure does not impose more restrictions than those implied by the "transparency" of $\Delta$.) Battigalli-Siniscalchi [10, 2007] and Battigalli-Prestipino [5, 2010] provide (distinct) formal treatments along these lines. They each get the $\Delta_a \times \Delta_b$-rationalizable strategy set, as an output.

**b. Two Characterization Theorems:** We have provided two characterizations of RCSBR—namely, the EFBRS solution concept and the Directed Rationalizability solution concept. While Proposition 5.1 shows that the two concepts are in a sense equivalent, we think that it is valuable to have both definitions on the table.

For the Directed Rationalizability definition: We already mentioned that there are times where the analyst understands that the context only imposes particular restrictions on players' first-order beliefs. In this case, the Directed Rationalizability procedure is useful. (See Section 8a.)

For the EFBRS definition: Often times, this definition is operationally "more convenient." We have seen that the EFBRS properties give us insight into behavior in games. (See Sections 6-7.) Moreover, there is a sense in which it may be "easier" to compute the solution concept, when beginning from the EFBRS definition vs. the Directed Rationalizability definition. In particular,

[12] This case is perhaps more relevant for applications of the theory of games with incomplete information, which is the focus of Battigalli-Siniscalchi[9]: An example of first-order restrictions "known to the analyst" is that hierarchies of initial beliefs about states of nature are derived from a given information structure.

to compute the concept according to the Directed Rationalizability definition, we must begin with each finite set of first-order beliefs and run the Directed Rationalizability procedure relative to each such set. (See Remark 5.1.) The set of all such finite sets has the cardinality of the continuum. On the other hand, to compute the concept according to the EFBRS definition, we begin with a subset of strategies, viz. $Q_a \times Q_b$, and verify conditions (i)-(iii) of Definition 5.1. There are a finite number of such sets $Q_a \times Q_b$.[13]

**c. Properties of EFBRS's:** Refer back to Sections 6-7. To analyze games of interest, we made use of the three properties of an EFBRS. Many of these arguments drew from Brandenburger-Friedenberg's [15, 2004] analysis of self-admissible sets: They began with properties of self-admissible sets (SAS's) and, analogously, used these properties to draw implications in terms of behavior in games.

While there is a close connection between the EFBRS properties and the SAS properties, there are also important points of difference. Indeed, the concepts are distinct. For an SAS, viz. $Q_a \times Q_b$, each $s_a \in Q_a$ must be admissible (i.e., not weakly dominated) in both the matrices $S_a \times S_b$ and $S_a \times Q_b$. For an EFBRS, we only require that each $s_a \in Q_a$ must be sequentially optimal under a CPS that strongly believes $Q_b$. If $s_a$ meets the former criterion, it meets the latter criterion, but the converse need not hold. So, in this sense, it is harder to meet the SAS criterion vs. the EFBRS criterion. On the other hand, SAS also has a maximality criterion, and it is easier to meet the SAS maximality criterion vs. the EFBRS maximality criterion.

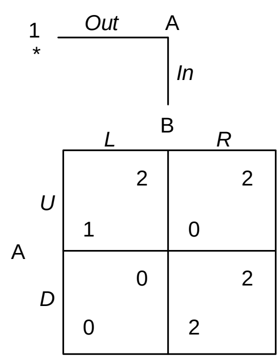


Figure 8.1

Putting these considerations together, we can have an EFBRS that is not an SAS, and an SAS

[13] But, we don't want to make too much of this point: Fix some $Q_a \times Q_b$. To check whether a particular strategy satisfies Definition 5.1, we must find some CPS satisfying conditions (i)-(iii). The set of all CPS's also has the cardinality of the continuum. In light of this, it may not be all that simple to check the EFBRS definition.

that is not an EFBRS. To see that an EFBRS need not be an SAS, refer to Figure 8.1. There, $\{Out\} \times \{Left, Right\}$ is an EFBRS, but the only SAS is $\{In\text{-}Down\} \times \{Right\}$. (Here, we use the admissibility criteria of SAS's.) To see that an SAS need not be an EFBRS, refer to Figure 1.4. There, $\{Out\} \times \{Left, Center\}$ is an SAS, but the only EFBRS is $\{In\text{-}Middle\} \times \{Center\}$. (Here, we use the fact that it is easier to meet the maximality criteria for SAS's vs. EFBRS's.)

**d. A Dominance Characterization of EFBRS's:** Fix a simultaneous move game and an associated type structure. Let us consider the conditions of "rationality and common belief of rationality." Here, we get, as an output, a best response set (Pearce [23, 1984]) $Q_a \times Q_b$. The definition of a best response set can be given both in terms of justifiability (i.e., each $s_a \in Q_a$ is optimal under a measure that assigns probability one to $Q_b$) and in terms of dominance (i.e., each $s_a \in Q_a$ is undominated in the matrix $S_a \times Q_b$). Likewise, if we consider the self-admissible set (Brandenburger-Friedenberg-Keisler [16, 2008]) concept, we can also provide a definition both in terms of justifiability and in terms of dominance.

Here, we have provided a justifiability definition of an EFBRS. On the game tree, the appropriate notion of dominance is "conditional dominance," i.e., undominated at each information set. (See Shimoji-Watson [27, 1998].) What about a conditional dominance characterization of an EFBRS? We don't know of such a characterization and leave it as an open question.

Let us comment on the essential difficulty in finding such a definition. It comes down to the maximality criterion. Definition 5.1 requires that we find some CPS $\mu_a$ that—in addition to satisfying conditions (i)-(ii)—also satisfies the requirement that, if $r_a$ is sequentially optimal under $\mu_a$, then $r_a \in Q_a$. Of course, strategies $r_a$ that are sequentially optimal under $\mu_a$ are conditionally undominated (see [27, 1998; Lemma 2]), but a conditionally undominated strategy need not be sequentially optimal under the given CPS $\mu_a$. Thus, we need a criterion to precisely say which conditionally undominated strategies $r_a$ must be included in $Q_a$.

There is a certain instance in which there is a clear criterion to say precisely which conditionally undominated strategies must be included in $Q_a$. Specifically, fix some $s_a \in Q_a$ and some $r_a$ that only allows information sets allowed by $s_a$. Here, we can build on the maximality criterion in [16, 2008], to give a precise criterion in terms of dominance. For simultaneous move games, any information set allowed by any strategy $r_a$ is also allowed by $s_a$. So, again, for simultaneous move games we can specify the appropriate maximality criterion. But, of course, for extensive-form games more generally, this condition need not be met. As such, more generally, the dominance criterion is not obvious—at least not to us.

We expand on these points in the Online Appendix.

**e. Existence of EFBRS's:** Note, the extensive-form rationalizable strategies form an EFBRS. (This is easily seen from Proposition 5.1, taking $\Delta_a \times \Delta_b$ to be the set of all CPS's.) As such, there exists a non-empty EFBRS. See Battigalli [3, 1997; Corollary 1].

**f. Two vs. Three Player Games:** Here, we have focused on two player games. The main results (Theorem 5.1 and Corollary 5.1) extend to the three player case, up to issues of correlation. Specifically, if we allow for correlated assessments in Definition 4.6, then we must also allow for correlated assessments in Definition 5.1. A similar statement holds for the case of independence—though, of course, care is needed in defining independence for CPS's. The central issue is that Charlie's belief about Bob should not change after Charlie learns information only about Ann. (It is easy to state this property for games with observable deviators. See, e.g., Battigalli [2, 1996]. Battigalli [2, 1996], Kohlberg-Reny [20, 1997], Stalnaker [29, 1998], and Swinkels [30, 1994] each address the definition, for more general games.)

Note, one additional issue that arises in the three player case: Should we require that Ann strongly believes "Bob and Charlie are rational"? Or should we instead require that Ann strongly believes "Bob is rational" and strongly believes "Charlie is rational"? Arguably, in the case of independence, we should require the latter.

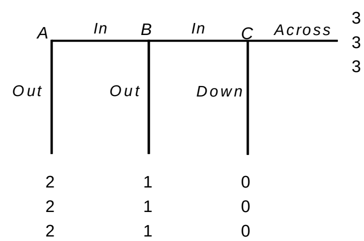


Figure 8.2

How does this affect our analysis of games? Amend Figure 6.3, so that it is a three-player game, as in Figure 8.2. Consider a state at which there is RCSBR in the sense explained above, and let's ask which strategies can be played. Of course, using rationality, Charlie must play *Across* (at this state). Note, now we require that a type of Bob strongly believe "Ann is rational" and also "Charlie is rational." So, conditional upon Bob's information set being reached, this type must maintain a hypothesis that Charlie is rational, and so that Charlie plays *Across*. In this case, there is a unique best response—namely, to play $In$. Turning to Ann, we see that under an RCSBR analysis she will choose $In$. So, we only get the backward induction outcome. (Battigalli-Siniscalchi [6, 1999] provide a "context free" epistemic analysis of this notion of independent rationalization. See Stalnaker [29, 1998] for a related idea.)

This example also shows that, in the case of independence, Proposition 7.1(ii) does not hold. Of course, if we instead consider the case of correlation and require that Bob strongly believe "Ann and Charlie are rational," then it may very well be the case that when Bob's node is reached he must forgo the hypothesis that Charlie is rational. Thus, in this case, we do have an analogue of Proposition 7.1(ii). Indeed, both parts (i)-(ii) of Proposition 7.1 hold for the case of correlation.

**g. Perfect Information Games:** In Section 7, we analyzed perfect information games and saw a connection between RCSBR and Nash outcomes. We already mentioned the connection to Brandenburger-Friedenberg's [15, 2010] SAS analysis. But there is another important connection to be made, namely to Ben Porath [11, 1997; Theorem 2 ].

The starting point in Ben Porath [11, 1997] is "rationality and common initial belief of rationality" (RCIBR). A type initially believes an event if it assigns probability one to the event at the initial node. (So, the type may initially believe an event, but not strongly believe the event.) RCIBR does not give a Nash outcome—for instance, in the Centipede Game of Figure 6.1, it would give $\{Out, Down\} \times \{Out, In\}$. However, Ben Porath goes on to show that, under an additional "grain of truth assumption," a Nash outcome does obtain (under a no ties condition). Interestingly, we may have a set of states consistent with RCSBR, where the grain of truth assumption does not obtain. There is a question if Ben Porath's conditions imply RCSBR—we do not know. Finally, we note that Ben Porath does not address a converse (under a no ties condition).

# Appendix A Self-Evident Events

Throughout the text, we informally argued that a type structure captures the idea that certain beliefs are "transparent" to the players. In this Appendix, we formalize the statement. The idea is that we will look at self-evident events and, in a precise sense clarified below, these events will correspond to the events that are "transparent" to players.

**I. Self-Evident Events.** Let us start with some preliminary definitions. Throughout, $(\Omega, \mathcal{B}(\Omega))$ is separable metrizable.

**Definition A1** *Fix a CPS* $\mu(\cdot|\cdot) : \Omega \times \mathcal{B}(\Omega) \times \mathcal{E} \to [0,1]$ *and an event* $E \in \mathcal{B}(\Omega)$. *Say* $\mu$ ***believes*** $E$ *if, for each* $F \in \mathcal{E}$, $\mu(E|F) = 1$.

In what follows, fix a game $\Gamma$ and a $\Gamma$-based type structure $\mathcal{T} = \langle S_a, S_b; \mathcal{S}_a, \mathcal{S}_b; T_a, T_b; \beta_a, \beta_b \rangle$.

**Definition A2** *Say a type* $t_a \in T_a$ ***believes*** $E_b \in \mathcal{B}(S_b \times T_b)$ *if* $\beta_a(t_a)$ *believes* $E_b$.

Given an event $E_b \in \mathcal{B}(S_b \times T_b)$, write $\mathrm{B}_a(E_b)$ for $S_a \times \{t_a \in T_a : t_a \text{ believes } E_b\}$. When $E_a \times E_b \in \mathcal{B}(S_a \times T_a \times S_b \times T_b)$, write $\mathrm{B}(E_a \times E_b) = \mathrm{B}_a(E_b) \times \mathrm{B}_b(E_a)$. Let us record two properties of belief, which will become useful as we proceed. (The proof of the first is straightforward, and so omitted.)

**Property A1 (Monotonicity)** *Fix events* $E_a \times E_b$ *and* $F_a \times F_b$ *in* $\mathcal{B}(S_a \times T_a \times S_b \times T_b)$. *If* $E_a \times E_b \subseteq F_a \times F_b$, *then* $\mathrm{B}(E_a \times E_b) \subseteq \mathrm{B}(F_a \times F_b)$.

**Property A2 (Conjunction)** *Fix a sequence of events* $E_a^1 \times E_b^1$, $E_a^2 \times E_b^2, \ldots$ *each in* $\mathcal{B}(S_a \times T_a \times S_b \times T_b)$. *Then* $\bigcap_m \mathrm{B}(E_a^m \times E_b^m) = \mathrm{B}(\bigcap_m (E_a^m \times E_b^m))$.

**Proof.** It is immediate from monotonicity that $\mathrm{B}(\bigcap_m (E_a^m \times E_b^m)) \subseteq \mathrm{B}(E_a^m \times E_b^m)$, for each $m$. As such, $\mathrm{B}(\bigcap_m (E_a^m \times E_b^m)) \subseteq \bigcap_m \mathrm{B}(E_a^m \times E_b^m)$. We now turn to the opposite inclusion, i.e., $\bigcap_m \mathrm{B}(E_a^m \times E_b^m) \subseteq \mathrm{B}(\bigcap_m (E_a^m \times E_b^m))$. Fix a type $t_a$ that believes each $E_b^m$, i.e., for each $h \in H_a$ and each $m$, $\beta_a(t_a)(E_b^m | S_b(h) \times T_b) = 1$. Define $F_b^m = \bigcap_{n=1}^m E_n^b$ and note that, for each $h$ and each (finite) $m$, $\beta_a(t_a)(F_b^m | S_b(h) \times T_b) = 1$. Then, for each $h \in H_a$, $\beta_a(t_a)(\bigcap_m F_b^m | S_b(h) \times T_b) = 1$. (This uses continuity of the probability measure $\beta_a(t_a)(\cdot | S_b(h) \times T_b)$.) Since, for each $h \in H_a$, $\beta_a(t_a)(\bigcap_m E_b^m | S_b(h) \times T_b) = \beta_a(t_a)(\bigcap_m F_b^m | S_b(h) \times T_b) = 1$, $t_a$ believes $\bigcap_m E_b^m$. ■

**Definition A3** *Say* $E_a \times E_b \in \mathcal{B}(S_a \times T_a \times S_b \times T_b)$ *is a* ***self-evident event (in*** $\mathcal{T}$***)*** *if* $E_a \times E_b \subseteq \mathrm{B}(E_a \times E_b)$.

We now proceed to relate self-evident events to those events that are "transparent." In particular, we will see that $E_a \times E_b$ is self-evident if and only if, at each state where $E_a \times E_b$ obtains, there is common belief that $E_a \times E_b$ obtains. More generally, a self-evident event always corresponds to the "transparency" of some (possibly different) event $F_a \times F_b$. For example, a self-evident event may reflect the idea that a certain event about "players' beliefs over strategies"—i.e., a certain event about "first-order beliefs"—is transparent. (In the main text, we referred to these events as representing "transparent restrictions" of first-order beliefs.)

Fix $E_a \times E_b \in \mathcal{B}(S_a \times T_a \times S_b \times T_b)$, and iterate the belief operator $\mathrm{B}(\cdot)$: $\mathrm{B}^0(E_a \times E_b) = E_a \times E_b$ and, for each $m \geq 0$, $\mathrm{B}^{m+1}(E_a \times E_b) = \mathrm{B}(\mathrm{B}^m(E_a \times E_b))$.

**Lemma A1** *Fix an event* $E_a \times E_b \in \mathcal{B}(S_a \times T_a \times S_b \times T_b)$. *The following are equivalent:*

*(i)* $E_a \times E_b$ *is self-evident (in* $\mathcal{T}$*);*

*(ii)* $E_a \times E_b = \bigcap_m \mathrm{B}^m(E_a \times E_b)$*;*

*(iii)* $E_a \times E_b = \bigcap_m \mathrm{B}^m(F_a \times F_b)$, *for some event* $F_a \times F_b \in \mathcal{B}(S_a \times T_a \times S_b \times T_b)$.

**Proof.** We show that (i) implies (ii). First note, for each event $E_a \times E_b$, $\mathrm{B}^0(E_a \times E_b) \cap \bigcap_{m \geq 1} \mathrm{B}^m(E_a \times E_b) \subseteq E_a \times E_b$. So, it suffices to show that, if $E_a \times E_b$ is a self-evident event, then $E_a \times E_b \subseteq \mathrm{B}^m(E_a \times E_b)$, for each $m \geq 1$. The case of $m = 1$ follows immediately from the fact that $E_a \times E_b$ is a self-evident event. Assume this is true for $m \geq 1$, i.e., $E_a \times E_b \subseteq \mathrm{B}^m(E_a \times E_b)$. Then, by monotonicity, $\mathrm{B}(E_a \times E_b) \subseteq \mathrm{B}(\mathrm{B}^m(E_a \times E_b))$. So, again using the fact that $E_a \times E_b$ is a self-evident event, we have that $E_a \times E_b \subseteq \mathrm{B}(E_a \times E_b) \subseteq \mathrm{B}^{m+1}(E_a \times E_b)$.

Next note that (ii) implies (iii), by taking $E_a = F_a$ and $E_b = F_b$. So, it suffices to show that (iii) implies (i).

For this, fix $E_a \times E_b$, $F_a \times F_b \in \mathcal{B}(S_a \times T_a \times S_b \times T_b)$ with $E_a \times E_b = \bigcap_{m \geq 0} \mathrm{B}^m (F_a \times F_b)$. Note that

$$
\begin{aligned}
E_a \times E_b &= \bigcap_{m\geq 0} \mathrm{B}^m (F_a \times F_b) \\
&= (F_a \times F_b) \cap (\bigcap_{m\geq 0} \mathrm{B} (\mathrm{B}^m (F_a \times F_b))) \\
&= (F_a \times F_b) \cap \mathrm{B}(\bigcap_{m\geq 0} \mathrm{B}^m (F_a \times F_b)) \\
&= (F_a \times F_b) \cap \mathrm{B}(E_a \times E_b),
\end{aligned}
$$

where the first and last lines use part (iii), the second line is by definition and the third line uses conjunction. It then follows that $E_a \times E_b \subseteq \mathrm{B}(E_a \times E_b)$ as required. ■

**II. Type Structures as Self-Evident Events.** We want to capture that a certain event is transparent to the players. We have argued that the idea is captured by the self-evident event concept. But, in the main text, we modelled the idea that an event is transparent by writing down some arbitrary type structure. How do the approaches relate? We will see that, in fact the approaches coincide. In particular, the self-evident events in a given type structure correspond to "smaller type structures." We first present the formal statement, and then review.

We will want to map one type structure into a second larger structure, and argue that, by doing so, we get a self-evident event. For this, it will be convenient to introduce some notation. Fix separable metrizable spaces $\Omega, \Phi$. Given a measurable map $f : \Omega \to \Phi$, write $\overline{f} : \mathcal{P}(\Omega) \to \mathcal{P}(\Phi)$, for the map where $\overline{f}(\varpi)$ is the image measure of $\varpi$ under $f$. Note, $\overline{f}$ is measurable. (See Kechris [18, 1995; Exercise 17.40].)

Now, consider two $\Gamma$-based type structures, namely $\mathcal{T} = \langle S_a, S_b; \mathcal{S}_a, \mathcal{S}_b; T_a, T_b; \beta_a, \beta_b \rangle$ and $\mathcal{T}^* = \langle S_a, S_b; \mathcal{S}_a^*, \mathcal{S}_b^*; T_a^*, T_b^*; \beta_a^*, \beta_b^* \rangle$. We will relate CPS's in structure $\mathcal{T}$ to CPS's in the structure $\mathcal{T}^*$. For this, it will be convenient to write $\mathrm{id}_a : S_a \to S_a$ and $\mathrm{id}_b : S_b \to S_b$ for the identity maps.

**Lemma A2** *Fix a measurable map $\tau_b : T_b \to T_b^*$ and a CPS $\mu_a \in \mathcal{C}(S_b \times T_b; \mathcal{S}_b)$. Define $\nu_a$ so that, for each $h \in H_a$, $(\overline{\mathrm{id}_b \times \tau_b})(\mu_a(\cdot|S_b(h) \times T_b)) = \nu_a(\cdot|S_b(h) \times T_b^*)$. Then $\nu_a \in \mathcal{C}(S_b \times T_b^*; \mathcal{S}_b^*)$.*

**Proof.** It is immediate that $\nu_a$ satisfies conditions (i)-(ii) of a CPS. For condition (iii), fix events $E^* \subseteq S_b(h) \times T_b^* \subseteq S_b(i) \times T_b^*$. Since a separable metrizable space is second countable, $(\mathrm{id}_b \times \tau_b)^{-1}(E^*) \in \mathcal{B}(S_b \times T_b)$. It follows that

$$
\begin{aligned}
\nu_a(E^*|S_b(i) \times T_b^*) &= \mu_a((\mathrm{id}_b \times \tau_b)^{-1}(E^*)|S_b(i) \times T_b) \\
&= \mu_a((\mathrm{id}_b \times \tau_b)^{-1}(E^*)|S_b(h) \times T_b) \times \mu_a(S_b(h) \times T_b|S_b(i) \times T_b) \\
&= \nu_a(E^*|S_b(h) \times T_b^*) \times \nu_a(S_b(h) \times T_b^*|S_b(i) \times T_b^*),
\end{aligned}
$$

as required. ■

Given CPS's $\mu_a$ and $\nu_a$ as in Lemma A2, say $\boldsymbol{\nu^a}$ **is the image CPS of** $\boldsymbol{\mu^a}$ **under** $\mathrm{id}^{\mathbf{b}} \times \boldsymbol{\tau^b}$. We write $\underline{(\mathrm{id}_b \times \tau_b)} : \mathcal{C}(S_b \times T_b; \mathcal{S}_b) \to \mathcal{C}(S_b \times T_b^*; \mathcal{S}_b^*)$ for the associated map, i.e., so that $\underline{(\mathrm{id}_b \times \tau_b)}(\mu_a)$ is the image CPS of $\mu_a$ under $\mathrm{id}_b \times \tau_b$. Note, for this, we make use of the fact that the map $\mathrm{id}_b \times \tau_b$ is measurable. (This follows from second countability.) Indeed, throughout, we will repeatedly make use of this fact.

**Definition A4** *Let $\tau_a : T_a \to T_a^*$ and $\tau_b : T_b \to T_b^*$ be measurable maps. Call $(\tau_a, \tau_b)$ a **type morphism** from $\mathcal{T}$ to $\mathcal{T}^*$ if $\underline{(\mathrm{id}_b \times \tau_b)} \circ \beta_a = \beta_a^* \circ \tau_a$ and $\underline{(\mathrm{id}_a \times \tau_a)} \circ \beta_b = \beta_b^* \circ \tau_b$.*

If $(\tau_a, \tau_b)$ is a type morphism, then each of $\tau_a$ and $\tau_b$ are hierarchy morphisms, i.e., mappings that preserve hierarchies of beliefs. See Section 3.1 in Battigalli-Siniscalchi [7, 1999].

Given separable metrizable spaces $\Omega, \Phi$, call a function $f : \Omega \to \Phi$ **bimeasurable** if it is measurable and, for each $E \in \mathcal{B}(\Omega)$, $f(E) \in \mathcal{B}(\Phi)$.

**Definition A5** *Call $(\tau_a, \tau_b)$ a **bimeasurable type morphism** if it is a type morphism and $\tau_a$ and $\tau_b$ are bimeasurable.*

We can now talk about the relationship between the self-evident event concept and the maps from one structure to a second larger structure.

**Lemma A3**

*(i) Fix $\Gamma$-based type structures $\mathcal{T} = \langle S_a, S_b; \mathcal{S}_a, \mathcal{S}_b; T_a, T_b; \beta_a, \beta_b \rangle$ and $\mathcal{T}^* = \langle S_a, S_b; \mathcal{S}_a^*, \mathcal{S}_b^*; T_a^*, T_b^*; \beta_a^*, \beta_b^* \rangle$. If $(\tau_a, \tau_b)$ is a bimeasurable type morphism from $\mathcal{T}$ to $\mathcal{T}^*$, then $S_a \times \tau_a(T_a) \times S_b \times \tau_b(T_b)$ is self-evident in $\mathcal{T}^*$.*

*(ii) Fix a $\Gamma$-based type structure $\mathcal{T}^* = \langle S_a, S_b; \mathcal{S}_a^*, \mathcal{S}_b^*; T_a^*, T_b^*; \beta_a^*, \beta_b^* \rangle$ and a self-evident event $S_a \times E_a^* \times S_b \times E_b^* \in \mathcal{B}(S_a \times T_a^* \times S_b \times T_b^*)$. Write $\tau_a : E_a^* \to T_a^*$ and $\tau_b : E_b^* \to T_b^*$ for the identity maps. Then there is a $\Gamma$-based type structure, viz. $\mathcal{T} = \langle S_a, S_b; \mathcal{S}_a^*, \mathcal{S}_b^*; E_a^*, E_b^*; \beta_a, \beta_b \rangle$, so that $(\tau_a, \tau_b)$ is a bimeasurable type morphism from $\mathcal{T}$ to $\mathcal{T}^*$.*

**Proof.** Begin with part (i). Fix a bimeasurable type morphism, viz. $(\tau_a, \tau_b)$. Since the maps $\tau_a$ and $\tau_b$ are bimeasurable, $S_a \times \tau_a(T_a) \times S_b \times \tau_b(T_b)$ is contained in $\mathcal{B}(S_a \times T_a^* \times S_b \times T_b^*)$. We proceed to show that $S_a \times \tau_a(T_a) \times S_b \times \tau_b(T_b) \subseteq \mathrm{B}(S_a \times \tau_a(T_a) \times S_b \times \tau_b(T_b))$. To show this, it suffices to show that, for each $\tau_a(t_a) \in \tau_a(T_a)$ and each information set $h \in H_a$,

$$\beta_a^*(\tau_a(t_a))(S_b \times \tau_b(T_b) | S_b(h) \times T_b^*) = 1.$$

(Again, bimeasurability guarantees that $S_b \times \tau_b(T_b)$ is Borel in $S_b \times T_b^*$.) But this follows from the definition of a type morphism, since

$$\beta_a^*(\tau_a(t_a))(S_b \times \tau_b(T_b) | S_b(h) \times T_b^*) = \beta_a(t_a)(S_b \times T_b | S_b(h) \times T_b) = 1.$$

Turn to part (ii). Take $T_a = E_a^*$, $T_b = E_b^*$, and endow these sets with the relative topology. (Recall that $S_b \times E_b^* \in \mathcal{B}(S_b \times T_b^*)$, so that $\mathcal{B}(S_b \times E_b^*) = \mathcal{B}(S_b \times T_b) \subseteq \mathcal{B}(S_b \times T_b^*)$.) For each $t_a \in E_a^*$, define $\beta_a(t_a)$ so that, for each $F_b \in \mathcal{B}(S_b \times T_b)$ and each $h \in H_a$,

$$\beta_a(t_a)(F_b|S_b(h) \times T_b) = \beta_a^*(t_a)(F_b|S_b(h) \times T_b^*).$$

Note, $\beta_a(t_a)$ defines a CPS with conditioning events $\mathcal{S}_a \otimes T_b$. To see this, recall that, for each $t_a \in T_a = E_a^*$ and for each $h \in H_a$,

$$\beta_a(t_a)(S_b(h) \times T_b|S_b(h) \times T_b) = \beta_a^*(t_a)(S_b(h) \times E_b^*|S_b(h) \times T_b^*) = 1,$$

where the first equality is by definition and the latter equality is by the fact that $t_a \in E_a^*$ and $S_a \times E_a^* \times S_b \times E_b^*$ is a self-evident event. This establishes condition (i) of a CPS. Conditions (ii)-(iii) are immediate from the construction.

We first show that $\mathcal{T} = \langle S_a, S_b; \mathcal{S}_a, \mathcal{S}_b; T_a, T_b; \beta_a, \beta_b \rangle$ is indeed a $\Gamma$-based type structure: It is immediate that $T_a$ and $T_b$ are separable metrizable. So, it suffices to show that $\beta_a$ and $\beta_b$ are measurable. We show this for $\beta_a$, and an analogous argument establishes the result for $\beta_b$.

To show that $\beta_a$ is measurable, it suffices to show that, for each information set $h \in H_a$, $t_a \mapsto \beta_a(t_a)(\cdot|S_b(h) \times T_b)$ is measurable. Note that $T_b = E_b^* \in \mathcal{B}(T_b^*)$. So, there is a homeomorphism

$$f : \mathcal{P}(S_b(h) \times T_b) \to \{\varpi \in \mathcal{P}(S_b(h) \times T_b^*) : \varpi(S_b(h) \times E_b^*) = 1\}.$$

(See Kechris [18, 1995; Exercise 17.28].) Now, fix an event $G_b \in \mathcal{B}(\mathcal{P}(S_b(h) \times T_b))$. Then, $f(G_b) \in \mathcal{B}(\mathcal{P}(S_b(h) \times T_b^*))$. By measurability of the map $\beta_a^*$, we have that

$$\{t_a^* \in T_a^* : \beta_a^*(t_a^*)(\cdot|S_b(h) \times T_b^*) \in f(G_b)\}$$

is Borel in $T_a^*$. But now notice that $f$ is such that

$$\{t_a \in T_a : \beta_a(t_a)(\cdot|S_b(h) \times T_b) \in G_b\} = \{t_a^* \in T_a^* : \beta_a^*(t_a^*)(\cdot|S_b(h) \times T_b^*) \in f(G_b)\} \cap E_a^*.$$

That is, $\{t_a \in T_a : \beta_a(t_a)(\cdot|S_b(h) \times T_b) \in G_b\}$ is an intersection of two measurable sets and so measurable. This establishes that $t_a \mapsto \beta_a(t_a)(\cdot|S_b(h) \times T_b)$ is measurable, as required.

Finally, consider the identity maps, viz. $\tau_a : T_a \to T_a^*$ and $\tau_b : T_b \to T_b^*$. Certainly, they are bimeasurable. We will show that $(\tau_a, \tau_b)$ is a type morphism. Fix a type $t_a \in E_a^*$ and we will show that $\beta_a^*(t_a)$ is the image CPS of $\beta_a(t_a)$ under $\mathrm{id}_b \times \tau_b$. Fix a Borel set $F_b^*$ in $S_b \times T_b^*$. For each $h \in H_a$,

$$\beta_a^*(t_a)(F_b^*|S_b(h) \times T_b^*) = \beta_a^*(t_a)(F_b^* \cap (S_b \times E_b^*)|S_b(h) \times T_b^*),$$

since $S_b \times E_a^* \times S_b \times E_b^*$ is a self-evident event. As such,

$$\begin{aligned}\beta_a^*(t_a)(F_b^*|S_b(h)\times T_b^*) &= \beta_a^*(t_a)(F_b^*\cap(S_b\times E_b^*)|S_b(h)\times T_b^*)\\ &= \beta_a(t_a)((\mathrm{id}_b\times\tau_b)^{-1}(F_b^*\cap(S_b\times E_b^*))|S_b(h)\times T_b)\\ &= \beta_a(t_a)((\mathrm{id}_b\times\tau_b)^{-1}(F_b^*)|S_b(h)\times T_b),\end{aligned}$$

as required. ■

Lemma A3 says that if there is a bimeasurable type morphism from $\mathcal{T}$ to $\mathcal{T}^*$, then $\mathcal{T}$ induces a self-evident event in $\mathcal{T}^*$. We now point out that we preserve RCSBR under the type morphism. Specifically, suppose there is a bimeasurable type morphism, viz. $(\tau_a, \tau_b)$, from $\mathcal{T}$ to $\mathcal{T}^*$. Let $E^*(\mathcal{T})$ be the self-evident event in $\mathcal{T}^*$ corresponding to $\mathcal{T}$, i.e., $E^*(\mathcal{T}) = S_a \times \tau_a(T_a) \times S_b \times \tau_b(T_b)$. Then, there is RCSBR at the state $(s_a, t_a, s_b, t_b)$ (in $\mathcal{T}$) if and only if there is rationality and common strong belief of "rationality and the self-evident event $E^*(\mathcal{T})$" at the state $(s_a, \tau_a(t_a), s_b, \tau_b(t_b)) \in E^*(\mathcal{T})$.

**Proposition A1** *Fix $\Gamma$-based structures $\mathcal{T}$ and $\mathcal{T}^*$, so that there is a bimeasurable type morphism, viz. $(\tau_a, \tau_b)$, from $\mathcal{T}$ to $\mathcal{T}^*$. Then, for each $m$,*

*(i) If $(s_a, t_a, s_b, t_b) \in \mathrm{CSB}^m(R_a \times R_b)$ then $(s_a, \tau_a(t_a), s_b, \tau_b(t_b)) \in \mathrm{CSB}^{*,m}((R_a^* \times R_b^*) \cap (S_a \times \tau_a(T_a) \times S_b \times \tau_b(T_b)))$.*

*(ii) If $(s_a, t_a^*, s_b, t_b^*) \in \mathrm{CSB}^{*,m}((R_a^* \times R_b^*)\cap(S_a\times\tau_a(T_a)\times S_b\times\tau_b(T_b)))$, then $(\tau_a)^{-1}(t_a^*), (\tau_b)^{-1}(t_b^*) \neq \emptyset$ and $\{s_a\} \times (\tau_a)^{-1}(t_a^*) \times \{s_b\} \times (\tau_b)^{-1}(t_b^*) \subseteq \mathrm{CSB}^m(R_a \times R_b)$.*

To prove Proposition A1, it will be useful to introduce some further notation. Fix two type structures $\mathcal{T}$ and $\mathcal{T}^*$, and a type morphism, viz. $(\tau_a, \tau_b)$, from $\mathcal{T}$ to $\mathcal{T}^*$. For the structure $\mathcal{T}$, write $R_a^0 = R_a$ and $R_b^0 = R_b$. Then, for each $m \geq 0$, set $R_a^{m+1} = R_a^m \cap \mathrm{SB}_a(R_b^m)$ and $R_b^{m+1} = R_b^m \cap \mathrm{SB}_b(R_a^m)$. It is easily verified that, for each $m$, $R_a^m \times R_b^m = \mathrm{CSB}^m(R_a \times R_b)$. For the structure $\mathcal{T}^*$, write $R_a^0 = R_a^* \cap [S_a \times \tau_a(T_a)]$ and $R_b^0 = R_b^* \cap [S_b \times \tau_b(T_b)]$. Then, for each $m \geq 0$, set $R_a^{*,m+1} = R_a^{*,m} \cap \mathrm{SB}_a^*(R_b^{*,m})$ and $R_b^{*,m+1} = R_b^{*,m} \cap \mathrm{SB}_b^*(R_a^{*,m})$. It is easily verified that, for each $m$, $R_a^{*,m} \times R_b^{*,m} = \mathrm{CSB}^{*,m}((R_a^* \times R_b^*) \cap (S_a \times \tau_a(T_a) \times S_b \times \tau_b(T_b)))$.

**Lemma A4** *Fix $\Gamma$-based structures $\mathcal{T}$ and $\mathcal{T}^*$, and a type morphism, viz. $(\tau_a, \tau_b)$, from $\mathcal{T}$ to $\mathcal{T}^*$. If $(s_a, t_a) \in R_a$, then $(s_a, \tau_a(t_a)) \in R_a^*$. Conversely, if $(s_a, t_a^*) \in R_a^*$, then $\{s_a\} \times (\tau_a)^{-1}(\{t_a^*\}) \subseteq R_a$.*

**Proof.** Fix some $t_a$ with $\tau_a(t_a) = t_a^*$. To show this result, it suffices to show that

$$\mathrm{marg}_{S_b(h)}\,\beta_a(t_a)(\cdot|S_b(h)\times T_b) = \mathrm{marg}_{S_b(h)}\,\beta_a^*(t_a^*)(\cdot|S_b(h)\times T_b^*),$$

for each $h \in H_a$. To see this, fix some information set $h \in H_a$ and some event $E_b \in \mathcal{B}(S_b(h))$. Then, by definition of a type morphism,

$$\beta_a^*(\tau_a(t_a))(E_b\times T_b^*|S_b(h)\times T_b^*) = \beta_a(t_a)(E_b\times(\tau_b)^{-1}(T_b^*)|S_b(h)\times T_b) = \beta_a(t_a)(E_b\times T_b|S_b(h)\times T_b),$$

as required. ■

**Proof of Proposition A1.** Given the characterization above, we will show that the following holds, for each $m$: (i) If $(s_a, t_a) \in R_a^m$, then $(s_a, \tau_a(t_a)) \in R_a^{*,m}$. (ii) If $(s_a, t_a^*) \in R_a^{*,m}$, then $\emptyset \neq \{s_a\} \times (\tau_a)^{-1}(\{t_a^*\}) \subseteq R_a^m$. And likewise for $b$.

We show this by induction on $m$. The case of $m = 0$ follows from Lemma A4. Assume the result holds for some $m \geq 0$ and we will show it also hold for $m+1$. Let us record two consequences of the induction hypothesis:

**Fact I** $R_b^{*,m} = (\mathrm{id} \times \tau_b)(R_b^m)$: Fix $(s_b, t_b^*) \in R_b^{*,m}$. Then, $(s_b, t_b^*) \in R_b^0$ and $t_b^* \in \tau_b(T_b)$. Fix some $t_b$ with $\tau_b(t_b) = t_b^*$. By the induction hypothesis, $(s_b, t_b) \in R_b^m$. So, $(s_b, t_b^*) = (s_b, \tau_b(t_b)) \in (\mathrm{id}_b \times \tau_b)(R_b^m)$, as required. The converse follows immediately from the induction hypothesis.

**Fact II** $R_b^m = (\mathrm{id} \times \tau_b)^{-1}((\mathrm{id} \times \tau_b)(R_b^m))$: Certainly, $R_b^m \subseteq (\mathrm{id} \times \tau_b)^{-1}((\mathrm{id} \times \tau_b)(R_b^m))$. Fix $(s_b, t_b) \in (\mathrm{id} \times \tau_b)^{-1}((\mathrm{id} \times \tau_b)(R_b^m))$. Then, using Fact I, $(s_b, \tau_b(t_b)) \in (\mathrm{id} \times \tau_b)(R_b^m) = R_b^{*,m}$. By part (ii) of the induction hypothesis, $(s_b, t_b) \in R_b^m$, as required.

We use these facts below.

First, fix $(s_a, t_a) \in R_a^{m+1}$. By the induction hypothesis, it suffices to show that $\tau_a(t_a)$ strongly believes $R_b^{*,m}$. First we show that $R_b^{*,m} \in \mathcal{B}(S_b \times T_b^*)$. To see this, use Fact I, i.e., $R_b^{*,m} = (\mathrm{id}_b \times \tau_b)(R_b^m)$. Since $t_a$ strongly believes $R_b^m$, it follows that $R_b^m \in \mathcal{B}(S_b \times T_b)$. Using the fact that $\mathrm{id}_b \times \tau_b$ is bimeasurable, we get that $R_b^{*,m}$ is indeed Borel.

Now, fix some information set $h \in H_a$ with $R_b^{*,m} \cap [S_b(h) \times T_b^*] \neq \emptyset$. Note that

$$\begin{aligned}
\beta_a^*(\tau_a(t_a))(R_b^{*,m}|S_b(h) \times T_b^*) &= \beta_a(t_a)((\mathrm{id}_b \times \tau_b)^{-1}(R_b^{*,m})|S_b(h) \times T_b) \\
&= \beta_a(t_a)(\{(s_b, t_b) : (s_b, \tau_b(t_b)) \in R_b^{*,m}\}|S_b(h) \times T_b) \\
&= \beta_a(t_a)(R_b^m|S_b(h) \times T_b)
\end{aligned}$$

where the first line follows from the definition of a type morphism, the second line is by definition, and the third line follows from the induction hypothesis. By part (ii) of the induction hypothesis, $R_b^m \cap [S_b(h) \times T_b] \neq \emptyset$. So, with the above and the fact that $t_a$ strongly believes $R_b^m$,

$$\beta_a^*(\tau_a(t_a))(R_b^{*,m}|S_b(h) \times T_b^*) = \beta_a(t_a)(R_b^m|S_b(h) \times T_b) = 1,$$

as required.

For the converse, fix $(s_a, t_a^*) \in R_a^{*,m+1}$ and some $t_a \in (\tau_a)^{-1}(t_a^*)$. By the induction hypothesis, it suffices to show that $t_a$ strongly believes $R_b^m$. Recall that $t_a^*$ strongly believes $R_b^{*,m}$ and so $R_b^{*,m} \in \mathcal{B}(S_b \times T_b^*)$. By Facts I-II, plus the observation that $\mathrm{id}_b \times \tau_b$ is measurable, $R_b^m = (\mathrm{id} \times \tau_b)^{-1}(R_b^{*,m})$ is Borel.

Now, fix an information set $h \in H_a$ with $R_b^m \cap [S_b(h) \times T_b] \neq \emptyset$. Note that

$$\begin{aligned} \beta_a(t_a)(R_b^m | S_b(h) \times T_b) &= \beta_a(t_a)((\mathrm{id}_b \times \tau_b)^{-1}((\mathrm{id}_b \times \tau_b)(R_b^m)) | S_b(h) \times T_b) \\ &= \beta_a^*(\tau_a(t_a))((\mathrm{id}_b \times \tau_b)(R_b^m) | S_b(h) \times T_b^*) \\ &= \beta_a^*(\tau_a(t_a))(R_b^{*,m} | S_b(h) \times T_b^*), \end{aligned}$$

where the first line follows from Fact II, the second line follows from the definition of a type morphism, and the last line follows from Fact I. By part (i) of the induction hypothesis, $R_b^{*,m} \cap [S_b(h) \times T_b^*] \neq \emptyset$. So, with the above and the fact that $\tau_a(t_a)$ strongly believes $R_b^{*,m}$,

$$\beta_a(t_a)(R_b^m | S_b(h) \times T_b) = \beta_a^*(\tau_a(t_a))(R_b^{*,m} | S_b(h) \times T_b^*) = 1,$$

as required. ■

**III. Self-Evident Events vs. Type Structures.** Let us review the approach taken here. We begin with a game $\Gamma$, and we will consider the canonical $\Gamma$-based type structure, as constructed in Battigalli-Siniscalchi [7, 1999]. Write $\mathcal{T}^* = \langle S_a, S_b; \mathcal{S}_a^*, \mathcal{S}_b^*; T_a^*, T_b^*; \beta_a^*, \beta_b^* \rangle$ for this structure. The details of the construction will not be relevant. Instead, we will make use of two properties. First, $\mathcal{T}^*$ is **complete**—that is, $\beta_a^*$ and $\beta_b^*$ are onto. (See Footnote 2.) Second, $\mathcal{T}^*$ is **terminal**—that is, for each $\Gamma$-based structure $\mathcal{T}$, there is a type morphism from $\mathcal{T}$ to $\mathcal{T}^*$.[14]

We can use Lemma A1 to generate the self-evident events in $\mathcal{T}^*$. To see this, return to the lady's choice convention. Let $F_a^* = S_a \times T_a^*$ and let $F_b^* = S_b \times \{t_b^* \in T_b^* : t_b^* \text{ believes } \{Up\} \times T_a^*\}$. Then, by Lemma A1(i)-(iii), we can find some $E_a^* \in \mathcal{B}(T_a^*)$ and $E_b^* \in \mathcal{B}(T_b^*)$, so that $S_a \times E_a^* \times S_b \times E_b^*$ is a self evident event, with $S_a \times E_a^* \times S_b \times E_b^* = \bigcap_m \mathrm{B}^m(F_a^* \times F_b^*)$. Certainly, each $t_b^* \in E_b^*$ believes $\{Up\} \times T_a^*$. Moreover, for each CPS $\mu_a \in \mathcal{C}(S_b \times T_b^*)$ (resp. $\mu_b \in \mathcal{C}(S_a \times T_a^*)$) that believes $S_b \times E_b^*$ (resp. $\{Up\} \times E_a^*$), there is a type $t_a^* \in T_a^*$ (resp. $t_b^* \in T_b^*$) with $\beta_a^*(t_a^*) = \mu_a$ (resp. $\beta_b^*(t_b^*) = \mu_b$). (Here, we use the fact that $\mathcal{T}^*$ is complete.) Indeed, the proof of Lemma A1(iii) gives that these types are in fact in $E_a^*$ (resp. $E_b^*$). So, by Lemma A3(ii), we can construct a type structure $\mathcal{T}$, as described in Section 1.1. Let $E^* = S_a \times E_a^* \times S_b \times E_b^*$ denote the self-evident event in $\mathcal{T}^*$ that corresponds to $\mathcal{T}$. Using Proposition A1, RCSBR within the constructed structure $\mathcal{T}$ corresponds to the event "rationality, $E^*$, and common strong belief of 'rationality and $E^*$'" within the canonical structure $\mathcal{T}^*$.

So, we see that we can indeed approach the question of a lady's choice convention, as we did in the main text. No need to work directly with self-evident events (in the canonical construction). Is this true more generally? Indeed, the answer is yes, and rests on the fact that the structure $\mathcal{T}^*$ is terminal. Because of this, we can find a type morphism from each $\Gamma$-based structure $\mathcal{T}$ to

[14] The terminology stems from Böge-Eisele [13, 1979]. Battigalli-Siniscalchi [7, 1999] show their construction is terminal, but they restrict attention to type structures with Polish type sets and continuous belief maps. The Online Appendix extends terminality to separable metrizable type sets and measurable belief maps.

the canonical Γ-based structure $\mathcal{T}^*$. When $\mathcal{T}$ satisfies certain conditions, the type morphism is bimeasurable. So, in this case, Lemma A3 and Proposition A1 give that the two approaches are equivalent. We will see that, in a certain sense, these conditions are "predominant." Let us explain.

We will impose two conditions on Γ-based type structure $\mathcal{T}$. First, the type sets $T_a$ and $T_b$ are standard Borel. Second, the type structure is countably uncountable—i.e., there are at most a countable number of hierarchies (of conditional beliefs), so that the set of types that induce that hierarchy is uncountable. In this case, the type morphism from $\mathcal{T}$ to $\mathcal{T}^*$ is bimeasurable. (Here we use Purves' Theorem [24, 1966], Proposition 3.3.7 in Srivastava [28], and the fact that the map from types to hierarchies is measurable. See the Online Appendix.) So, to the extent that we can restrict attention to standard Borel countably uncountable structures, Lemma A3 and Proposition A1 give that the two approaches are indeed equivalent.

Now: Can we indeed restrict attention to standard Borel and countably uncountable structures? Yes. Begin with Theorem 5.1. Note that the type structure constructed in part (ii) is finite, so certainly satisfies these conditions. Next turn to Lemma A3. Note that the type structure constructed in part (ii) also satisfies these conditions. The fact that $T_a$ and $T_b$ are standard Borel follows from the fact that they are Borel subsets of a Polish space. The fact that the constructed structure is countably uncountable follows from the fact that $T_a$ and $T_b$ are finite.

# Appendix B Proofs for Section 4

**Proof of Property 4.1.** Fix an event $F \in \mathcal{E}$ with $F \cap \bigcap_m E_m \neq \emptyset$. Then $F \cap E_m \neq \emptyset$ for all $m$. So, for each $m$, $\mu(E_m|F) = 1$. (This is because $\mu$ strongly believes each $E_m$.) But then $\mu(\bigcap_m E_m|F) = 1$. ■

**Proof of Property 4.2.** Fix an event $F \in \mathcal{E}$ with $F \cap \operatorname{proj}_{\Omega_1} E \neq \emptyset$. Then $(F \times \Omega_2) \cap E \neq \emptyset$. Since, by assumption, $\operatorname{proj}_{\Omega_1} E$ is Borel, $\operatorname{marg}_{\Omega_1} \mu \left(\operatorname{proj}_{\Omega_1} E|F\right)$ is well defined. Since $\mu$ strongly believes $E$, $\mu(E|F \times \Omega_2) = 1$. Then $(\operatorname{marg}_{\Omega_1} \mu)\left(\operatorname{proj}_{\Omega_1} E|F\right) = 1$, as required. ■

# Appendix C RCSBR and Directed Rationalizability

In the text, we argued that, for each epistemic type structure, there is a set of first-order beliefs $\Delta$ so that the projection of the RCSBR set is the $\Delta$-rationalizable strategy set. The purpose of this appendix is to illustrate that this set of first-order beliefs may not correspond to the set of all first-order beliefs allowed by the epistemic type structure.

Figure B1 is a game of Battle of the Sexes preceded by an observed "money burning" move by Bob. (See Ben Porath-Dekel [12, 1992].) Here, Ann and Bob are playing a BoS game. However, prior to the game, Bob has the option of burning ($B$) or not burning ($NB$) \$2.

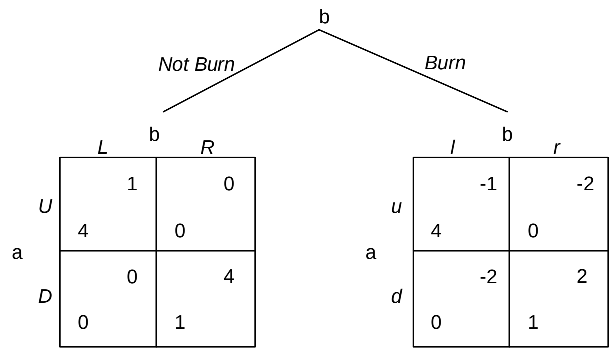


Figure B1

Suppose society has formed a modified version of the lady's choice convention. Now, there are no restrictions on players' first-order beliefs. (So, in particular, there are types of Bob that think Ann does not go for her best payoff.) But, there is a restriction on Ann's second-order beliefs. Specifically, conditional upon observing so-called "normal" behavior – i.e., a decision to not burn – Ann thinks that Bob thinks she goes for her best payoff and chooses $Up$. There is no restriction on Ann's second-order belief conditional upon observing "strange" behavior – i.e., upon observing a decision to burn. Likewise, there are no restrictions on Bob's second-order beliefs. Etc.

We can model this modified version of the lady's choice convention by a type structure $\langle S_a, S_b; \mathcal{S}_a, \mathcal{S}_b; T_a, T_b; \beta_a, \beta_b \rangle$ based on the game in Figure B1. Now, $\beta_b$ is onto but $\beta_a$ is not. Formally: Write $[Up]_a$ for the event "Ann plays $Up$, if Bob does not burn," i.e., $[Up]_a = \{Up\text{-}down, Up\text{-}up\} \times T_a$, and write $[NB]_b$ for the event "Bob does not burn," i.e., $[NB]_b = \{NB\text{-}Left, NB\text{-}Right\} \times T_b$. Let $U_b$ be the set of types $t_b \in T_b$ with $\beta_b(t_b)([Up]_a | S_a \times T_a) = 1$, i.e., the set of types of Bob that assign probability one to the event "Ann plays $Up$, when Bob chooses not to burn." Then, for each type $t_a \in T_a$,

$$\beta_a(t_a)(S_b \times U_b | [NB]_b) = 1,$$

i.e., conditional upon Bob choosing not to burn, each type of Ann assigns probability one to the event that "Bob believes that 'Ann plays $Up$, when Bob does not burn.'" For any belief $\mu_a$ of Ann with $\mu_a(S_b \times U_b | [NB]_b) = 1$, there is a type $t_a$ so that $\beta_a(t_a) = \mu_a$. (See Appendix A in [**?**, 2009] on how to construct such a type structure.)

The set of first-order beliefs induced by this type structure is $\Delta = \mathcal{C}(S_b) \times \mathcal{C}(S_a)$. The $\Delta$-rationalizable set is $\{Down\text{-}down\} \times \{NB\text{-}Right\}$. (This is also the set of extensive form rationalizable strategies.) It is obtained as follows: On round one, the strategy $B\text{-}left$ is dominated by

$NB\text{-}Left$, but all other strategies (of both players) are optimal under some CPS. It follows that

$$S_a^{\Delta,1} \times S_b^{\Delta,1} = S_a \times \{NB\text{-}Left, NB\text{-}Right, B\text{-}right\}.$$

But now note that the choice of *up* by Ann cannot be optimal under any CPS that strongly believes $\{NB\text{-}Left, NB\text{-}Right, B\text{-}right\}$. (If a CPS strongly believes $\{NB\text{-}Left, NB\text{-}Right, B\text{-}right\}$, then conditional upon $Burn$ being played, the CPS must assign probability one to $right$, in which case *up* is not a best response.) So,

$$S_a^{\Delta,2} \times S_b^{\Delta,2} = \{Up\text{-}down, Down\text{-}down\} \times S_b^{\Delta,1}.$$

Turning to Bob, if a CPS strongly believes $\{Up\text{-}down, Down\text{-}down\}$, then $B\text{-}right$ yields an expected payoff of 2 and $NB\text{-}Left$ yields an expected payoff of at most one. So,

$$S_a^{\Delta,3} \times S_b^{\Delta,3} = S_a^{\Delta,2} \times \{NB\text{-}Right, B\text{-}right\}.$$

Now, if a CPS strongly believes $\{NB\text{-}Right, B\text{-}right\}$, $Down\text{-}down$ is the only sequentially optimal strategy, so

$$S_a^{\Delta,4} \times S_b^{\Delta,4} = \{Down\text{-}down\} \times S_b^{\Delta,3}.$$

Finally, if a CPS strongly believes $\{Down\text{-}down\}$, $NB\text{-}Right$ is the only sequentially optimal strategy, so

$$S_a^{\Delta,5} \times S_b^{\Delta,5} = \{Down\text{-}down\} \times \{NB\text{-}Right\}.$$

But, the projection of event RCSBR onto $S_a \times S_b$ is $\{Up\text{-}down\} \times \{B\text{-}right\}$. It is obtained as follows. On round one, for each belief about the strategies of the other player, there is a type that holds that belief. So, here too,

$$\mathrm{proj}_{S_a} R_a^1 \times \mathrm{proj}_{S_b} R_b^1 = S_a \times \{NB\text{-}Left, NB\text{-}Right, B\text{-}right\}.$$

Now, consider a type $t_a$ that strongly believes $R_b^1$. Recall, conditional upon Bob choosing not to burn, each type of Ann assigns probability one to the event that "Bob believes that 'Ann plays $Up$, when Bob does not burn.'" So, if $t_a$ strongly believes $R_b^1$, it must assign zero probability to $\{NB\text{-}Right\} \times T_b$. For such a type $t_a$, $(Down\text{-}down, t_a)$ is irrational. So,

$$\mathrm{proj}_{S_a} R_a^2 \times \mathrm{proj}_{S_b} R_b^2 = \{Up\text{-}down\} \times \mathrm{proj}_{S_b} R_b^1.$$

But now, if $(s_b, t_b)$ is rational and $t_b$ strongly believes $R_a^2$, then $s_b = B\text{-}right$, and so

$$\mathrm{proj}_{S_a} R_a^3 \times \mathrm{proj}_{S_b} R_b^3 = \{Up\text{-}down\} \times \{B\text{-}right\}.$$

Why the difference between the two approaches? We began with an epistemic structure and used the structure itself to form the set of first-order beliefs $\Delta = \mathcal{C}(S_b) \times \mathcal{C}(S_a)$. (So, for each $\mu_a \in \Delta_a = \mathcal{C}(S_b)$ there is type $t_a \in T_a$ such that the marginal of $\beta_a(t_a)$ on $S_b$ is $\mu_a$; and likewise for $b$.) With this set of first-order beliefs, the strategies that survive one round of $\Delta$-rationalizability are exactly the strategies that are consistent with rationality. But, on the next round, we lose the equivalence: If $\beta_a(t_a)$ strongly believes $R_b^1$, then the marginal of $\beta_a(t_a)$ must strongly believe $S_b^{\Delta,1} = \text{proj}_{S_b} R_b^1$. (Here, we use the marginalization property of strong belief.) Thus $\text{proj}_{S_a} R_a^2 \subseteq S_a^{\Delta,2}$. But, the converse does not hold. We have *Down-down* $\in S_a^{\Delta,2}$, but *Down-down* $\notin \text{proj}_{S_a} R_a^2$. The reason is that, conditional upon Bob choosing $NB$, each $\beta_a(t_a)$ assigns probability one to the event "Bob assigns probability one to $[Up]_a$." So, if Bob does not burn, Ann can only maintain a hypothesis that Bob is rational, if she assigns probability one to Bob's playing $NB\text{-}Left$, in which case the choice *Down* is not a best response. With this, $S_a^{\Delta,2} = \{$*Up-down*, *Down-down*$\}$ and $\text{proj}_{S_a} R_a^2 = \{$*Up-down*$\}$. As a result, $S_b^{\Delta,3} = \{$*NB-Right*, *B-right*$\}$ and $\text{proj}_{S_b} R_b^3 = \{$*B-right*$\}$. It follows that $S_a^{\Delta,4} = \{$*Down-down*$\}$, despite the fact that $\text{proj}_{S_a} R_a^4 = \{$*Up-down*$\}$. The key to this last step is that *Up-down* is optimal under a CPS that strongly believes $\text{proj}_{S_b} R_b^3 \subsetneq S_b^{\Delta,3}$, but not optimal under a CPS that strongly believes $S_b^{\Delta,3}$. This can occur because strong belief fails a monotonicity requirement.

# Appendix D Proofs for Section 6

We begin by showing that, for the finitely repeated Prisoner's Dilemma, any EFBRS results in the *Defect-Defect* path of play. To show this, we will need to make use of certain properties of EFBRS's. We will again make use of these properties in Appendix E. We begin with the best response property.

**Definition D1** *Say $Q_a \times Q_b \subseteq S_a \times S_b$ satisfies the* ***best response property*** *if, for each $s_a \in Q_a$ there is a CPS $\mu_a \in \mathcal{C}(S_b)$, so that $s_a \in \rho_a(\mu_a)$ and $\mu_a$ strongly believes $Q_b$. And similarly for $b$.*

An EFBRS satisfies the best response property. But the converse need not hold, i.e., $Q_a \times Q_b$ may satisfy the best response property, but fail to be an EFBRS because it violates the maximality condition. (See the example in Section 1.5.)

Let us introduce some notation, to relate the whole game to its parts. Fix a game $\Gamma$ and a subgame $\Sigma$. Write $H_a^{\Sigma}$ for the set of information sets that are contained in $\Sigma$. We will abuse notation and write $S_a(\Sigma)$ for the set of strategies of $\Gamma$ that allow $\Sigma$. We also write $S_a^{\Sigma} = \prod_{h \in H_a^{\Sigma}} C_a(h)$ for the set of strategies of $a$ in the subgame $\Sigma$. Note, each strategy $s_a^{\Sigma} \in S_a^{\Sigma}$ can be viewed as the projection of a strategy $s_a \in S_a(\Sigma)$ into $S_a^{\Sigma}$. Given a set $E_a \subseteq S_a$, write $E_a^{\Sigma}$ for the set of strategies $s_a^{\Sigma} \in S_a^{\Sigma}$ so that there is some $s_a \in E_a \cap S_a(\Sigma)$ whose projection into $S_a^{\Sigma}$ is $s_a^{\Sigma}$. We will write $\pi_a^{\Sigma}$ and $\pi_b^{\Sigma}$ for the payoff functions associated with the subtree $\Sigma$. So, if $(s_a, s_b)$ allows $\Sigma$, then $\pi^{\Sigma}\left(s_a^{\Sigma}, s_b^{\Sigma}\right) = \pi(s_a, s_b)$.

**Lemma D1** *Fix a game $\Gamma$ and a subgame $\Sigma$. If $Q_a \times Q_b$ satisfies the best response property for the game $\Gamma$, then $Q_a^\Sigma \times Q_b^\Sigma$ satisfies the best response property for the subgame $\Sigma$.*

**Proof.** If $Q_a^\Sigma \times Q_b^\Sigma = \emptyset$ (if no profile in $Q_a \times Q_b$ allows $\Sigma$), then it is immediate that $Q_a^\Sigma \times Q_b^\Sigma$ satisfies the best response property. So, we will suppose $Q_a^\Sigma \times Q_b^\Sigma \neq \emptyset$.

Fix a strategy $s_a^\Sigma \in Q_a^\Sigma$. Then there exists a strategy $s_a \in Q_a \cap S_a(\Sigma)$ whose projection into $\prod_{h \in H_a^\Sigma} C_a(h)$ is $s_a^\Sigma$. Since $s_a \in Q_a$, we can find a CPS $\mu_a \in \mathcal{C}(S_b)$ so that $s_a \in \rho_a(\mu_a)$ and $\mu_a$ strongly believes $Q_b$.

Let $\mathcal{S}_b^\Sigma$ be the set of all $S_b^\Sigma(h)$ for $h \in H_a^\Sigma$. Define $\nu_a^\Sigma(\cdot|\cdot) : \mathcal{B}(S_b^\Sigma) \times \mathcal{S}_b^\Sigma \to [0,1]$ so that, for each event $E_b \subset S_b$ and each $S_b^\Sigma(h) \in \mathcal{S}_b^\Sigma$, $\nu_a^\Sigma(E_b^\Sigma | S_b^\Sigma(h)) = \mu_a(E_b | S_b(h))$. It is readily verified that $\nu_a^\Sigma$ is indeed a CPS on $(S_b^\Sigma, \mathcal{S}_b^\Sigma)$.

Since $s_a$ allows $\Sigma$ and $s_a$ is sequentially optimal under $\mu_a$, it follows that $s_a^\Sigma$ is sequentially optimal under $\nu_a^\Sigma$. Fix some $S_b^\Sigma(h) \in \mathcal{S}_b^\Sigma$. If $Q_b^\Sigma \cap S_b^\Sigma(h) \neq \emptyset$, then $Q_b \cap S_b(h) \neq \emptyset$. So, in this case, $\nu_a^\Sigma(Q_b^\Sigma | S_b^\Sigma(h)) = \mu_a(Q_b | S_b(h)) = 1$. This establishes that $\nu_a^\Sigma$ strongly believes $Q_b^\Sigma$.

Interchanging $a$ and $b$ establishes the result. ■

We use Lemma D1 to show:

**Lemma D2** *Consider the $N$-repeated Prisoner's Dilemma, as given in Figure 6.2. If $Q_a \times Q_b$ satisfies the best response property for this game, then each strategy profile in $Q_a \times Q_b$ results in the Defect-Defect path.*

**Proof.** The proof very closely follows the proof of Example 3.2 in Brandenburger-Friedenberg [15, 2004]. It is by induction on $N$. For $N = 1$, the result is immediate. Assume the result holds for some $N$ and we will show it holds for $N + 1$.

Consider some $Q_a \times Q_b$ of the $N+1$ repeated Prisoner's Dilemma that satisfies the best response property. Suppose, there is a strategy $s_a \in Q_a$ that Cooperates in the first period. Fix a strategy $s_b \in Q_b$. If $s_b$ plays Cooperate (resp. Defect) in the first period, Ann gets $c$ (resp. $e$) in the first period. By Lemma D1 and the induction hypothesis, Ann gets a payoff of zero, in periods $2, \ldots, N$. So, for each $s_b$ in $Q_b$, $\pi_a(s_a, s_b) = c$ if $s_b$ plays Cooperate in the first period, and $\pi_a(s_a, s_b) = e$ if $s_b$ plays Defect in the first period.

Now, instead consider the strategy $r_a$ that plays Defect in every period, irrespective of the history. Again, fix a strategy $s_b \in Q_b$. If $s_b$ plays Cooperate in the first period, then $\pi_a(r_a, s_b) \geq d$ and, if $s_b \in Q_b$ plays Defect in the first period, then $\pi_a(r_a, s_b) \geq 0$.

Putting the above together: Under any CPS that strongly believes $Q_b$, we must have that $r_a$ is a strictly better response than $s_a \in Q_a$, at the first information set. But this contradicts $Q_a \times Q_b$ satisfying the best response property. ■

**Corollary D1** *Consider the $N$-repeated Prisoner's Dilemma, as given in Figure 6.2. If $Q_a \times Q_b$ is an EFBRS, then each strategy profile in $Q_a \times Q_b$ results in the Defect-Defect path.*

Now we turn to Proposition 6.1. We will show the result for a somewhat more general set of games – games where, in a sense, the information structure is determined by the subgames.

**Definition D2** *Fix a game* $\Gamma$. *Say a subgame* $\Sigma$ *is* ***sufficient*** *for an information set* $h \in H$ *if* $h$ *is contained in* $\Sigma$ *and the set of strategy profiles that allow* $\Sigma$ *is exactly* $S_a(h) \times S_b(h)$.

Notice, there may be two subgames, viz. $\Sigma$ and $\bar{\Sigma}$, that are sufficient for $h$.[15] If so, either $\Sigma$ is a subgame of $\bar{\Sigma}$ or $\bar{\Sigma}$ is a subgame of $\Sigma$. When there are two subgames that are sufficient for $h$, we will in typically be interested in the **last subgame** $\Sigma$ **sufficient for** $h$ – i.e., so that no proper subgame of $\Sigma$ is sufficient for $h$.

Also, notice that there may be no subgame that is sufficient for an information set $h$. Refer to the game in Figure 7.3. There, the only subgame is the entire game. But this subgame is not sufficient for the information set, viz. $h$, at which Bob moves. To see this, notice that the strategy $s_a = Out$ (trivially) allows the subgame, but does not allow $h$.

**Definition D3** *Say a game* $\Gamma$ *is* ***determined by its subgames*** *if, for each information set* $h \in H$, *there is a subgame* $\Sigma$ *that is sufficient for* $h$.

The game in Figure 7.3 is not determined by its subgame; as we have seen, there is no subgame that is sufficient for the information set at which Bob moves. Below, we will characterize Definition D3 in terms of primitives of the game (as opposed to a condition about strategies).

Throughout, we restrict attention to a game $\Gamma$ determined by its subgames. Fix a pure-strategy SPE, viz. $(s_a, s_b)$, of $\Gamma$. Construct maps $f_a : H \to S_a$ and $f_b : H \to S_b$ that depend on this SPE. To do so, fix some $h \in H$, and let $\Sigma$ be the last subgame sufficient for $h$. Write $x$ for the root of subgame $\Sigma$ (which may be $\Gamma$ itself). If $\Sigma = \Gamma$, set $f_a(h) = s_a$. If $\Sigma$ is a proper subtree of $\Gamma$, then we can write $x = (c^1, ..., c^K)$. In this case, let $f_a(h)$ be the strategy that (i) chooses $c_a^1$ at $\{\phi\}$, (ii) chooses $c_a^k$ at an information set that contains $(c^1, ..., c^{k-1})$, i.e., an initial segment of $(c^1, ..., c^K)$, and (iii) makes the same choice as $s_a$ at all other information sets. So, if $s_a$ allows $h$, then $f_a(h) = s_a$. Also, $f_a(h)$ is well-defined and allows $h$ precisely because $\Gamma$ is determined by its subgames. (Again, refer to the game in Figure 7.3, and take $h$ to be the information set at which Bob moves. Consider the SPE $(s_a, s_b) = (Out, Right)$. Then, $f_a(h) = Out$, which precludes $h$.)

Write $S(h)$ for the set of strategy profiles that allow an information set $h$. In games determined by their subgames, there is a natural order on sets of the form $S(h)$, for $h \in H$. Specifically, for any pair of information sets $h$ and $i$ (in $H$), either $S(h) \subseteq S(i)$, $S(i) \subseteq S(h)$, or $S(h) \cap S(i) = \emptyset$.[16] To see this, let $\Sigma_h$ (resp. $\Sigma_i$) be sufficient for $h$ (resp. $i$). We have that, either $\Sigma_h$ is a subgame of $\Sigma_i$, $\Sigma_i$ is a subgame of $\Sigma_h$, or they are disjoint subgames. With this, the order follows from the definition of sufficiency. If $S(h) \subseteq S(i)$, say $h$ **follows** $i$. Say $h$ and $i$ are **ordered** if either $h$ follows $i$ or $i$ follows $h$. Say $h$ and $i$ are **unordered** otherwise, i.e., if $S(h) \cap S(i) = \emptyset$.

[15] This may happen if there is a node $x$ where no player is active, i.e., $C_a(x)$ and $C_b(x)$ are singletons.

[16] Note, in all perfect recall games, whenever $h, i \in H_a$, either $S(h) \subseteq S(i)$, $S(i) \subseteq S(h)$, or $S(h) \cap S(i) = \emptyset$. Here, we have analogous statement, when $h \in H_a$ and $i \in H_b$.

Let us record a couple of facts, to be used below. The first is immediate.

**Lemma D3** *Fix a game $\Gamma$ that is determined by its subgames. Also fix some SPE, viz. $(s_a, s_b)$. Construct $(f_a, f_b)$ as above. If $f_a(h)$ allows $i$ and either $h$ and $i$ are unordered or $i$ follows $h$, then $f_a(i) = f_a(h)$.*

The next result is immediate from the definition of an SPE.

**Lemma D4** *Fix a game $\Gamma$ that is determined by its subgames and some SPE $(s_a, s_b)$. For each $h \in H_a$,*

$$\pi_a(f_a(h), f_b(h)) \geq \pi_a(r_a, f_b(h)) \quad \textit{for all } r_a \in S_a(h).$$

The next result holds quite generally. Again, its proof is immediate.

**Lemma D5** *Fix some $\mu_a \in \mathcal{C}(S_b)$. If $s_a \in \rho_a(\mu_a)$, then $[s_a] \subseteq \rho_a(\mu_a)$.*

The notion of a player being "decisive" for $(z, z')$ (Definition 6.2) was stated in the main text for games with observable actions. In order to have an appropriate generalization of Proposition 6.1 we have to extend this definition to games with imperfectly observable actions.

**Definition D4** *Fix two distinct terminal nodes $z = (x, c^1, \ldots, c^K)$ and $z' = (x, d^1, \ldots, d^L)$. Say* **a *is decisive for*** $(\mathbf{z}, \mathbf{z'})$ *if the following holds:*

*(i)* $c_a^1 \neq d_a^1$,

*(ii)* $c_b^1 = d_b^1$, *and*

*(iii) if $(x, c^1, \ldots, c^k)$ and $(x, d^1, \ldots, d^l)$ are in the same information set, then $c_b^{k+1} = d_b^{l+1}$.*

The idea is that $a$ is decisive for $(z, z') = ((x, c^1, \ldots, c^K), (x, d^1, \ldots, d^L))$ if $a$ is the only player that determines which of the two terminal histories occurs. So, $a$ moves at the last common predecessor of $z$ and $z'$, viz. $x$, and makes distinct choices at this node, i.e., $c_a^1 \neq d_a^1$. But, $b$'s choice along this path does not determine which of $z$ vs. $z'$ occurs. So, $b$ makes the same choice whenever he cannot observe $a$'s choice amongst $c_a^1$ vs. $d_a^1$.

**Remark D1** *If the game has observable actions, then $a$ is decisive for $(z, z') = ((x, c^1, \ldots, c^K), (x, d^1, \ldots, d^L))$ if and only if $c_a^1 \neq d_a^1$, and $c_b^1 = d_b^1$.*

Given Definition D4, the no relevant ties (NRT) property can be stated as in the main text, and we have the following:

**Proposition D1** *Fix a game $\Gamma$ that is determined by its subgames, and a pure-strategy SPE, viz. $(s_a, s_b)$.*

*(i) There is an EFBRS, viz. $Q_a \times Q_b$, so that $[s_a] \times [s_b] \subseteq Q_a \times Q_b$.*

*(ii) If $\Gamma$ satisfies NRT, then $[s_a] \times [s_b]$ is an EFBRS.*

**Proof.** Fix a pure-strategy SPE, viz. $(s_a, s_b)$. Construct maps $f_a : H \to S_a$ and $f_b : H \to S_b$, as above. We use these maps to construct CPS's $\mu_a \in \mathcal{C}(S_b)$ and $\mu_b \in \mathcal{C}(S_a)$. Specifically, set $\mu_a(f_b(h) | S_b(h)) = 1$ for each $h \in H_a$. And likewise for $a$ and $b$ interchanged.

To see that $\mu_a$ is indeed a CPS: It is immediate that $\mu_a$ satisfies conditions (i)-(ii) of Definition 3.1. For condition (iii), fix information sets $h, i \in H_a$ so that $S_b(i) \subseteq S_b(h)$. If $f_b(h) \in S_b(i)$, then $f_b(i) = f_b(h)$. (Lemma D3.) So, for each event $E \subseteq S_b(i)$,

$$\mu_a(E|S_b(h)) = \mu_a(E|S_b(i)) \times 1 = \mu_a(E|S_b(i))\, \mu_a(S_b(i) | S_b(h)).$$

If $f_b(h) \notin S_b(i)$, then for each event $E \subseteq S_b(i)$,

$$\mu_a(E|S_b(h)) = 0 = \mu_a(E|S_b(i)) \times 0 = \mu_a(E|S_b(i))\, \mu_a(S_b(h) | S_b(i)),$$

as required. And, likewise, for $b$.

Now, let $Q_a = \rho_a(\mu_a)$, i.e., the set of all strategies $r_a$ that are sequentially optimal under $\mu_a$. And, likewise, set $Q_b = \rho_b(\mu_b)$. We will show that $Q_a \times Q_b$ is an EFBRS.

Fix some $r_a \in Q_a$. We will show that $r_a$ and $\mu_a$ jointly satisfy conditions (i)-(iii) of an EFBRS. In fact, it is immediate that Conditions (i) and (iii) are satisfied. So, we will show condition (ii), i.e., that $\mu_a$ strongly believes $Q_b$.

Fix an information set $h \in H_a$ with $Q_b \cap S_b(h) \neq \emptyset$. We will show that $f_b(h) \in Q_b$, so that $\mu_a(Q_b|S_b(h)) = 1$. To show that $f_b(h) \in Q_b$, it suffices to show that, for each information set $i \in H_b$ allowed by $f_b(h)$,

$$\pi_b(f_a(i), f_b(h)) \geq \pi_b(f_a(i), r_b) \quad \text{for all } r_b \in S_b(i). \tag{D1}$$

Note, if either $i$ follows $h$ or $h$ and $i$ are unordered, then $f_b(h) = f_b(i)$. In either case, we can apply Lemma D4 to the information set $i$ and get the desired result. So, we focus on the case where $h$ follows $i$.

Take $S(h) \subseteq S(i)$**.** Since $Q_b \cap S_b(h) \neq \emptyset$, there is a strategy $r_b \in Q_b \cap S_b(h)$. For this strategy $r_b$, we have that $\pi_b(f_a(i), r_b) \geq \pi_b(f_a(i), f_b(h))$, because $r_b$ is sequentially optimal under $\mu_b$, $\mu_b(f_a(i) | S_a(i)) = 1$, and $f_b(h) \in S_b(h) \subseteq S_b(i)$. We will show that $\pi_b(f_a(i), r_b) = \pi_b(f_a(i), f_b(h))$, establishing Equation D1.

Suppose, contra hypothesis, that $\pi_b(f_a(i), r_b) > \pi_b(f_a(i), f_b(h))$. Consider the information set $j$, so that the last common predecessor of $(f_a(i), r_b)$ and $(f_a(i), f_b(h))$ is contained in $j$. Now, use the fact that $r_b$ and $f_b(h)$ both allow $h$, to get that either $j$ follows $h$ or $j$ and $h$ are unordered. In these cases, we have that $\pi_b(f_a(j), f_b(h)) \geq \pi_b(f_a(j), r_b)$. (This was established in the previous paragraph.) But now notice that, since either $j$ follows $h$ or $j$ and $h$ are unordered, we also have that either $j$ follows $i$ or $j$ and $i$ are unordered. In either case, using the fact that $f_a(i)$ allows $j$, we have $f_a(i) = f_a(j)$. (Lemma D3.) So, putting the above facts together,

$$\begin{aligned}\pi_b(f_a(i), f_b(h)) &= \pi_b(f_a(j), f_b(h)) \\ &\geq \pi_b(f_a(j), r_b) \\ &= \pi_b(f_a(i), r_b) \geq \pi_b(f_a(i), f_b(h)).\end{aligned}$$

But this contradicts the assumption that $\pi_b(f_a(i), r_b) > \pi_b(f_a(i), f_b(h))$.

We have established that $Q_a \times Q_b = \rho_a(\mu_a) \times \rho_b(\mu_b)$ is an EFBRS. By construction, $(s_a, s_b) \in \rho_a(\mu_a) \times \rho_b(\mu_b)$. So, using Lemma D5, $[s_a] \times [s_b] \subseteq Q_a \times Q_b$. Now, suppose the game tree has NRT. We will show that, if $(r_a, r_b) \in Q_a \times Q_b$, then $(r_a, r_b) \in [s_a] \times [s_b]$.

Fix some strategy $r_a \notin [s_a]$. Then, there exists some $r_b \in S_b$ with $\zeta(s_a, r_b) \neq \zeta(r_a, r_b)$. Consider the last common predecessor of $\zeta(s_a, r_b)$ and $\zeta(r_a, r_b)$, viz. $x$, and let $h$ be the information set that contains this node. Then, there exists $(c^1, ..., c^K)$ and $(d^1, ..., d^L)$ so that $\zeta(s_a, r_b) = (x, c^1, ..., c^K)$, $\zeta(r_a, r_b)) = (x, d^1, ..., d^L)$. Clearly, $c_a^1 = s_a(h) \neq r_a(h) = d_a^1$ and $c_b^k = r_b(h') = d_b^l$ whenever $(x, c^1, ..., c^{k-1}), (x, d^1, ..., d^L) \in h' \in H_b$. So, $a$ is decisive for $(\zeta(s_a, r_b), \zeta(r_a, r_b))$.

Now, by the analysis above, we have that $\pi_a(s_a, f_b(h)) \geq \pi_a(r_a, f_b(h))$.. NRT says that, in fact, $\pi_a(s_a, f_b(h)) > \pi_a(r_a, f_b(h))$. This implies that $r_a \notin Q_a$, as required. ■

**Lemma D6** *If $\Gamma$ has observable actions, then $\Gamma$ is determined by its subgames.*

**Proof.** Fix an information set $h$. Since $\Gamma$ has observable actions, $h = \{x\}$ for some node/history $x$. Now, consider a node $y$ that follows $x$. Then, by observable actions, $y$ is contained in the information set $\{y\}$. It follows that there is a subgame whose initial node is $x$, written $\Sigma$. Moreover, the set of strategies that allow $\Sigma$ is exactly $S_a(h) \times S_b(h)$. So, $\Gamma$ is determined by its subgames. ■

**Proof of Proposition 6.1.** Immediate from Proposition D1 and Lemma D6. ■

Finally, we return to characterize the condition that $\Gamma$ is determined by its subgames, in terms of primitives of the game tree alone (i.e., without reference to strategies). For this, we will need some notation: Given a set of nodes, viz. $\{x^1, \ldots, x^K\}$, write $\text{lcp}(\{x^1, \ldots, x^K\})$ for the last common predecessor of these nodes.

**Lemma D7** *A game $\Gamma$ is determined by its subgames if and only if, for each information set $h \in H$, the following holds:*

*(i) the last common predecessor of nodes in $h$, viz.* $\text{lcp}(h)$, *is the root of a subgame, and*

*(ii) the set of terminal nodes allowed by* $\text{lcp}(h)$ *is exactly the set of terminal nodes allowed by $h$.*

**Proof.** Fix a game $\Gamma$ and an information set $h$. First observe that if conditions (i)-(ii) are satisfied for $h$, then there must be some subgame sufficient for $h$. To see this claim, take $\Sigma$ to be the subgame whose root is $\text{lcp}(h)$. (Here we use condition (i).) Fix a strategy profile $(s_a, s_b)$ that allows $\text{lcp}(h)$.

Note that the terminal node $\zeta(s_a, s_b)$ is also allowed by $h$. (Here we use condition (ii).) So, $(s_a, s_b)$ must allow $h$. This establishes that $\Sigma$ is sufficient for $h$.

Now, we suppose that there is some subgame that is sufficient for $h$, viz. $\Sigma$. We will show that conditions (i)-(ii) must be satisfied. For this, we will make use of the fact that $\Sigma$ must contain $\text{lcp}(h)$.

First, we show condition (i). Suppose, contra hypothesis, $\text{lcp}(h)$ is contained in a non-singleton information set – i.e., there is some $x \neq \text{lcp}(h)$ so that $x$ and $\text{lcp}(h)$ are contained in the same information set. Then, $\text{lcp}(\{\text{lcp}(h), x\})$ is also contained in $\Sigma$. Moreover, there is some player who is active at $\text{lcp}(\{\text{lcp}(h), x\})$. This player has a strategy that allows $\Sigma$, but not $h$. This, contradicts the presumption that $\Sigma$ is sufficient for $h$.

Next is condition (ii). To see this, observe that the set of terminal nodes allowed by $h$ is contained in the set of terminal nodes allowed by $\text{lcp}(h)$. Fix a terminal node, viz. $z$, allowed by $\text{lcp}(h)$. Then, $z$ is also allowed by $\Sigma$ (since $\text{lcp}(h)$ is contained in the subtree $\Sigma$). So, there is a strategy profile $(s_a, s_b)$ that allows $\Sigma$ with $\zeta(s_a, s_b) = z$. Since $\Sigma$ is sufficient for $h$, $(s_a, s_b)$ allows $h$. It follows that $z$ is allowed by $h$, as required. ■

# Appendix E Proofs for Section 7

In this appendix, we prove Proposition 7.1. We also further discuss the gap between parts (i)-(ii) of the Proposition.

**I. Proof of Proposition 7.1(i):** This will follow immediately from the following Lemma.

**Lemma E1** *Fix a perfect-information game satisfying SPC. If $Q_a \times Q_b$ satisfies the best response property, then each $(s_a, s_b) \in Q_a \times Q_b$ is outcome equivalent to a Nash Equilibrium.*

The proof of this Lemma closely follows the proof of Proposition 6.1a in Brandenburger-Friedenberg [15, 2010]. It is by induction on the length of the tree. Specifically, fix a game $\Gamma$ and a subgame $\Sigma$. The induction hypothesis states that if a set satisfies the best response property on $\Sigma$ then it is outcome equivalent to some Nash equilibrium. We saw that, if a set $Q_a \times Q_b$ satisfies the best response property on $\Gamma$, it also satisfies the best response property on the subgame $\Sigma$. (This was Lemma D1 in Appendix D.) So, if we fix a set that satisfies the best response property on the whole tree, then, by the induction hypothesis, it is outcome equivalent to a Nash equilibrium on each reached subgame. The proof uses this fact to construct a pure strategy Nash equilibrium on the whole tree, that is outcome equivalent to each profile in $Q_a \times Q_b$.

Let us begin filling in the dots.

**Definition E1** *Call $Q_a \times Q_a \subseteq S_a \times S_b$ a **constant set** if, for each $(s_a, s_b)$, $(r_a, r_b) \in Q_a \times Q_b$, $\pi(s_a, s_b) = \pi(r_a, r_b)$.*

**Lemma E2** *Fix a perfect-information game satisfying SPC. If $Q_a \times Q_b$ satisfies the best response property, then $Q_a \times Q_b$ is a constant set.*

**Proof.** The proof is by induction on the length of the tree.

First, fix a tree of length one and suppose Ann moves at the initial node. Then Bob's strategy set is a singleton. So, if $Q_a \times Q_b$ satisfies the best response property, then Ann is indifferent between each $(s_a, s_b)$ and $(r_a, s_b)$ in $Q_a \times Q_b$. By SPC, each profile in $Q_a \times Q_b$ is outcome equivalent.

Assume the result holds for any tree of length $l$ or less. Fix a tree of of length $l+1$ and a set $Q_a \times Q_b$ satisfying the best response property. Suppose Ann moves at the initial node, and can choose amongst nodes $n_1, \ldots, n_K$. Each $n_k$ can be identified with an information set and each is associated with a subgame $\Sigma = k$.

In particular, fix some subgame $k$ with $Q_a^k \times Q_b^k \neq \emptyset$. Then $Q_a^k \times Q_b^k$ satisfies the best response property for the subgame $k$. (This is Lemma D1.) So, by the induction hypothesis, $\pi^k\left(s_a^k, s_b^k\right) = \pi^k\left(r_a^k, r_b^k\right)$, for each $\left(s_a^k, s_b^k\right)$ and $\left(r_a^k, r_b^k\right) \in Q_a^k \times Q_b^k$. Now, note that, for each $s_b \in Q_b$, $s_b^k \in Q_b^k$. (Here, we use the fact that Ann moves at the initial node.) Thus, given two strategies $s_a, r_a \in Q_a \cap S_a(\Sigma)$ and $s_b, r_b \in Q_b$, we have that $\pi(s_a, s_b) = \pi(r_a, r_b)$.

Now, fix some $(s_a, s_b), (r_a, r_b) \in Q_a \times Q_b$, where $s_a \in S_a(k)$ and $r_a \in S_a(j)$. We have already established that $\pi(s_a, s_b) = \pi(r_a, r_b)$, for $k = j$. Suppose $k \neq j$. Since $s_a \in Q_a$, $s_a$ is sequentially optimal under some $\mu_a(\cdot|\cdot)$ that strongly believes $Q_b$. So, in particular, $s_a$ is optimal under $\mu_a(\cdot|S_b)$ with $\mu_a(Q_b|S_b) = 1$. With this,

$$\begin{aligned}
\pi_a(s_a, s_b) &= \textstyle\sum_{q_b \in Q_b} \pi_a(s_a, q_b)\, \mu_a(q_b|S_b) \\
&\geq \textstyle\sum_{q_b \in Q_b} \pi_a(r_a, q_b)\, \mu_a(q_b|S_b) \\
&= \pi_a(r_a, r_b).
\end{aligned}$$

(The first equality follows from the fact that, for each $q_b \in Q_b$, $\pi_a(s_a, s_b) = \pi_a(s_a, q_b)$. This is a consequence of the last line in the preceding paragraph. Likewise, for the last equality.) By an analogous argument, $\pi_a(r_a, r_b) \geq \pi_a(s_a, s_b)$. So, $\pi_a(r_a, r_b) = \pi_a(s_a, s_b)$. Using the single payoff condition, $\pi_b(r_a, r_b) = \pi_b(s_a, s_b)$. ■

**Proof of Lemma E1.** The proof is by induction on the length of the tree.

First, fix a tree of length one and suppose Ann moves at the initial node. Then Bob's strategy set is a singleton. The result follows from the fact that each $s_a \in Q_a$ is sequentially optimal under a CPS.

Now assume the result holds for any tree of length $l$ or less. Suppose Ann moves at the initial node, and can choose among nodes $n^1, \ldots, n^K$. Each $n^k$ can be identified with an information set and each is associated with a subgame $\Sigma = k$.

Fix some $(s_a, s_b) \in Q_a \times Q_b$ and suppose $s_a \in S_a(1)$. Note, $Q_a^1 \times Q_b^1$ satisfies the best response property (Lemma D1). So, by the induction hypothesis, there is a Nash equilibrium of subgame 1, viz. $\left(r_a^1, r_b^1\right)$, so that $\pi\left(s_a^1, s_b^1\right) = \pi\left(r_a^1, r_b^1\right)$. Consider a strategy $r_a \in S_a(1)$ so that the projection of $r_a$ onto $\prod_{h \in H_a^1} C_a(h)$ is $r_a^1$. We need to show that we can choose $r_b^2, \ldots, r_b^K \in \times_{k=2}^K S_b^k$ so that, for each $q_a \in Q_a$ and associated $q_a^k \in S_a^k$, $\pi_a\left(r_a^1, r_b^1\right) \geq \pi_a\left(q_a^k, r_b^k\right)$. The profile $\left(r_a, \left(r_b^1, r_b^2, \ldots, r_b^K\right)\right)$ will then be a Nash Equilibrium of the game.

Since $s_a \in Q_a$, there exists a CPS and an associated measure $\mu_a(\cdot|S_b)$ so that

$$\sum\nolimits_{s_b \in S_b} \left[\pi_a(s_a, s_b) - \pi_a(q_a, s_b)\right] \mu_a(s_b|S_b) \geq 0,$$

for all $q_a \in S_a$. Fix $k$ from $2, \ldots, K$. Using Lemma E2,

$$\pi_a\left(r_a^1, r_b^1\right) = \pi_a\left(s_a^1, s_b^1\right) \geq \sum\nolimits_{s_b^k \in S_k^b} \pi_a\left(q_a^k, s_b^k\right)\left(\mathrm{marg}_{S_b^k}\, \mu(\cdot|S_b)\right)\left(s_b^k\right),$$

for any $q_a^k \in S_a^k$. Letting $(\overline{q}_a^k, \overline{q}_b^k) \in \arg\max_{S_a^k} \min_{S_b^k} \pi_a(\cdot,\cdot)$, we have in particular

$$\pi_a\left(r_a^1, r_b^1\right) \geq \sum\nolimits_{s_b^k \in S_k^b} \pi_a\left(\overline{q}_a^k, s_b^k\right)\left(\mathrm{marg}_{S_k^b}\, \mu(\cdot|S_b)\right)\left(s_b^k\right).$$

But $\pi_a(\overline{q}_a^k, q_b^k) \geq \pi_a(\overline{q}_a^k, \overline{q}_b^k)$ for any $q_b^k \in S_b^k$, by definition. So

$$\pi_a\left(r_a^1, r_b^1\right) \geq \sum\nolimits_{s_b^k \in S_k^b} \pi_a(\overline{q}_a^k, \overline{q}_b^k)(\mathrm{marg}_{S_b^k}\, \mu(\cdot|S_b))\left(s_b^k\right) = \pi_a(\overline{q}_a^k, \overline{q}_b^k).$$

Set $(\underline{q}_a^k, \underline{q}_b^k) \in \arg\min_{S_b^k} \max_{S_a^k} \pi_a(\cdot,\cdot)$. By the Minimax Theorem for PI games (see, e.g., Ben Porath [11, 1997]), $\pi_a(\overline{q}_a^k, \overline{q}_b^k) = \pi_a(\underline{q}_a^k, \underline{q}_b^k)$. It follows that $\pi_a(r_a^1, r_b^1) \geq \pi_a\left(\overline{q}_a^k, \overline{q}_b^k\right) = \pi_a(\underline{q}_a^k, \underline{q}_b^k)$. But $\pi_a(\underline{q}_a^k, \underline{q}_b^k) \geq \pi_a(q_a^k, \underline{q}_b^k)$ for any $q_a^k \in S_a^k$, by definition. So $\pi_a(r_a^1, r_b^1) \geq \pi_a(q_a^k, \underline{q}_b^k)$, for each $q_a^k \in S_a^k$. Setting each $r_b^k = \underline{q}_b^k$ gives the desired profile. ■

**II. Proof of Proposition 7.1(ii):** Let us give the idea of the proof. We will start with a set $Q_a \times Q_b = \{(s_a, s_b)\}$, where $(s_a, s_b)$ is a pure Nash equilibrium in sequentially justifiable strategies. This set will satisfy the best response property. (See Lemma E4 below.) In particular, the set $Q_a$ is associated with a single CPS $\mu_a$, satisfying the conditions of the best response property. We will look at the set $P_a$ of all strategies $r_a$ that are sequentially optimal under $\mu_a$. We use the fact that $\mu_a$ strongly believes $Q_b$ (so assigns probability 1 to $s_b$ at the initial information set) to get that Ann is indifferent between all outcomes associated with $P_a \times Q_b$. Indeed, by NRT, these strategy profiles must reach the same terminal node. Likewise, we define $P_b$ and, using standard properties of a PI game tree, we get that all strategies in $P_a \times P_b$ reach the same terminal node.

So, what have we done: We began with a set $Q_a \times Q_b$ and we expanded it to a set $P_a \times P_b$, with (i) $Q_a \times Q_b \subseteq P_a \times P_b$, (ii) all the profiles in $P_a \times P_b$ reach the same terminal node, and (iii) there is a CPS $\mu_a$ (resp. $\mu_b$) that strongly believes $Q_b$ (resp. $Q_a$) and such that $P_a$ (resp. $P_b$) is the set of strategies that are sequentially optimal under $\mu_a(\cdot|\cdot)$ (resp. $\mu_b(\cdot|\cdot)$). We would have succeeded in constructing an EFBRS if the CPS $\mu_a$ (resp. $\mu_b$) strongly believed $P_b$ (resp. $P_a$) instead of $Q_b$ (resp. $Q_a$). The key will be that we can similarly expand $P_a \times P_b$ so that the new set satisfies similar properties. Since the game is finite, eventually, the expanded set must coincide with the original set—that is, condition (i) must hold with equality. This gives the desired result.

Now we turn to the proof. First, we give a technical Lemma.

**Lemma E3** *Fix some* $(\Omega, \mathcal{E})$ *where* $\Omega$ *is finite. Let* $\mu(\cdot|\cdot)$ *be a CPS on* $(\Omega, \mathcal{E})$ *and let* $\varpi$ *be a measure on* $\Omega$*. Construct* $\nu(\cdot|\cdot) : \mathcal{B}(\Omega) \times \mathcal{E} \to [0,1]$ *as follows: If* $F \in \mathcal{E}$ *with* $\mathrm{Supp}\,\varpi \cap F \neq \emptyset$ *then* $\nu(\cdot|F) = \varpi(\cdot|F)$*. Otherwise,* $\nu(\cdot|F) = \mu(\cdot|F)$*. Then* $\nu(\cdot|\cdot)$ *is a CPS.*

**Proof.** Let $\mu$, $\varpi$, and $\nu$ be as in the statement of the Lemma. Conditions (i)-(ii) of a CPS are immediate. Turn to condition (iii). For this, fix $E \in \mathcal{B}(\Omega)$ and $F, G \in \mathcal{E}$ with $E \subseteq F \subseteq G$.

First suppose that $\mathrm{Supp}\,\varpi \cap F \neq \emptyset$. Then

$$\begin{aligned} \nu(E|G) &= \frac{\varpi(E)}{\varpi(G)} \\ &= \frac{\varpi(E)}{\varpi(F)} \frac{\varpi(F)}{\varpi(G)} = \nu(E|F)\,\nu(F|G), \end{aligned}$$

where the first equality makes use of the fact that $E \subseteq G$ and the last makes use of the fact that $E \subseteq F$ and $F \subseteq G$. Next suppose that $\mathrm{Supp}\,\varpi \cap G = \emptyset$. Then $\mathrm{Supp}\,\varpi \cap F = \emptyset$, so that

$$\begin{aligned} \nu(E|G) &= \mu(E|G) \\ &= \mu(E|F)\,\mu(F|G) = \nu(E|F)\,\nu(F|G), \end{aligned}$$

as required. Finally, suppose that $\mathrm{Supp}\,\varpi \cap F = \emptyset$ but $\mathrm{Supp}\,\varpi \cap G \neq \emptyset$. Then

$$0 \leq \nu(E|G) \leq \nu(F|G) = \varpi(F|G) = 0,$$

where the last equality follows from the fact that $\mathrm{Supp}\,\varpi \cap F = \emptyset$. Then

$$\begin{aligned} \nu(E|G) &= 0 \\ &= \mu(E|F)\,\varpi(F|G) = \nu(E|F)\,\nu(F|G), \end{aligned}$$

as required. ■

**Lemma E4** *Let* $(s_a, s_b)$ *be a Nash equilibrium in sequentially justifiable strategies. Then* $\{(s_a, s_b)\}$ *satisfies the best response property.*

**Proof.** Let $(s_a, s_b)$ be a Nash equilibrium in sequentially justifiable strategies. Then there exists a CPS $\mu_a(\cdot|\cdot)$ so that $s_a$ is sequentially optimal under $\mu_a(\cdot|\cdot)$. Construct a CPS $\nu_b(\cdot|\cdot)$ so that $\nu_b(s_b|S_b(h)) = 1$ if $s_b \in S_b(h)$, and $\nu_b(\cdot|S_b(h)) = \mu_a(\cdot|S_b(h))$ otherwise. By Lemma E3, $\nu_b(\cdot|\cdot)$ is a CPS. It is immediate from the construction that $s_a$ is sequentially optimal under $\nu_b(\cdot|\cdot)$ and $\nu_b(\cdot|\cdot)$ strongly believes $\{s_b\}$. And, similarly, with $a$ and $b$ reversed. ■

**Definition E2** *Fix a constant set $Q_a \times Q_a \subseteq S_a \times S_b$. Call $P_a \times P_a \subseteq S_a \times S_b$* ***an expansion of*** *$Q_a \times Q_b$ if there exists a CPS $\mu_a \in \mathcal{C}(S_b)$ so that:*

*(i)* $Q_a \subseteq P_a = \rho_a(\mu_a)$,

*(ii)* $\mu_a$ *strongly believes* $Q_b$, *and*

*(iii) if $r_a$ is optimal under $\mu_a(\cdot|S_b)$ then $\pi_a(r_a, s_b) = \pi_a(s_a, s_b)$ for all $(s_a, s_b) \in Q_a \times Q_b$.*

*And, likewise, with $a$ and $b$ reversed.*

Notice, we only define an expansion of a set $Q_a \times Q_b$, if $Q_a \times Q_b$ is a constant set. Also, note, if $P_a \times P_b$ is an expansion of $Q_a \times Q_b$ then there are CPS's $\mu_a$ and $\mu_b$ satisfying conditions (i)-(iii) of Definition E2. We will refer to these as **the associated CPS's**.

**Lemma E5** *Fix a PI game satisfying NRT. Suppose $P_a \times P_b$ is an expansion of $Q_a \times Q_b$ and fix associated CPS's $\mu_a$ and $\mu_b$. Let $X_a$ be the set of strategies that are optimal under $\mu_a(\cdot|S_b)$. And, likewise, define $X_b$. Then $X_a \times X_b$ is a constant set.*

**Proof.** Since $P_a \times P_b$ is an expansion of $Q_a \times Q_b$, $Q_a \times Q_b$ is a constant set. (This is by definition.) It follows from condition (iii) of Definition E2 that $X_a \times Q_b$ and $Q_a \times X_b$ are constant sets. Then, using NRT, each profile in $X_a \times Q_b$ reaches the same terminal node. And likewise for $Q_a \times X_b$. In fact, the terminal node reached by $X_a \times Q_b$ and $Q_a \times X_b$ must be the same one, since $(X_a \times Q_b) \cap (Q_a \times X_b) = (Q_a \times Q_b)$. Now fix a profile $(s_a, r_b) \in (X_a \backslash Q_a) \times (X_b \backslash Q_b)$. Note there is a profile $(s_a, s_b) \in (X_a \backslash Q_a) \times Q_b$ and a profile $(r_a, r_b) \in Q_a \times (X_b \backslash Q_b)$. These profiles reach the same terminal node and so $(s_a, r_b)$ must also reach that terminal node. This establishes that $X_a \times X_b$ is a constant set. ■

**Corollary E1** *Fix a PI game satisfying NRT. If $P_a \times P_b$ is an expansion of some $Q_a \times Q_b$, then $P_a \times P_b$ is constant.*

The next result is standard, and so the proof is omitted.

**Lemma E6** *Fix a measure $\varpi_a \in \mathcal{P}(S_b)$ so that $s_a$ is optimal under $\varpi_a$ given $S_a$. Then, for any information set $h$ with $s_a \in S_a(h)$ and $\varpi_a(S_b(h)) > 0$, $s_a$ is optimal under $\varpi_a(\cdot|S_b(h))$ given $S_a(h)$.*

Given a measure $\varpi \in \mathcal{P}(\Omega)$, we write $\mathrm{Supp}\,\varpi$ for the support of the measure.

**Lemma E7** *Fix a PI game satisfying NRT. If $P_a \times P_b$ is an expansion of $Q_a \times Q_b$, then there exists some $W_a \times W_b$ that is an expansion of $P_a \times P_b$.*

**Proof.** Begin with the fact that $P_a \times P_b$ is an expansion of $Q_a \times Q_b$, and choose an associated CPS $\mu_a$ (resp. $\mu_b$) satisfying the conditions of Definition E2. Let $X_a$ (resp. $X_b$) be the set of strategies that are optimal under $\mu_a(\cdot|S_b)$ (resp. $\mu_b(\cdot|S_a)$). By Lemma E5, $X_a \times X_b$ is a constant set.

Construct a measure $\varpi_a \in \mathcal{P}(S_b)$ as follows: Begin with a measure $\overline{\varpi}_a$ with $\operatorname{Supp}\overline{\varpi}_a = P_b$. Construct $\varpi_b$ so that, for each $r_b \in P_b$,

$$\varpi_a(r_b) = (1-\varepsilon)\,\mu_a(r_b|S_b) + \varepsilon\overline{\varpi}_a(r_b),$$

where $\varepsilon \in (0,1)$. Note that $\mu_a$ strongly believes $Q_b \subseteq P_b$, $\operatorname{Supp}\mu_a(\cdot|S_b) \subseteq P_b$. With this and the fact that $\operatorname{Supp}\overline{\varpi} = P_b$, we have $\operatorname{Supp}\varpi_a = P_b$. Using the fact that $X_a \times P_b$ is a constant set, $\pi_a(s_a, \varpi_a) = \pi_a(r_a, \varpi_a)$ for all $s_a, r_a \in X_a$. Moreover, when $\varepsilon$ is sufficiently small, $\pi_a(s_a, \varpi_a) > \pi_a(r_a, \varpi_a)$ for all $s_a \in X_a$ and $r_a \in S_a \backslash X_a$. So we can choose $\varpi_a$ so that $s_a$ is optimal under $\varpi_a$ if and only if $s_a \in X_a$.

Now construct a CPS $\nu_a \in \mathcal{C}(S_b)$ as follows: If $P_b \cap S_b(h) \neq \emptyset$, let $\nu_a(\cdot|S_b(h)) = \varpi_a(\cdot|S_b(h))$. (This is well defined since, in this case, $\varpi_a(S_b(h)) > 0$.) If $P_b \cap S_b(h) = \emptyset$, let $\nu_a(\cdot|S_b(h)) = \mu_a(\cdot|S_b(h))$. Lemma E3 establishes that $\nu_a(\cdot|\cdot)$ is a CPS. Construct a measure $\varpi_b \in \mathcal{P}(S_a)$ and a CPS $\nu_b \in \mathcal{C}(S_a)$ analogously.

Take $W_a = \rho_a(\nu_a)$ and $W_b = \rho_b(\nu_b)$. We will show that $W_a \times W_b$ is an expansion of $P_a \times P_b$.

Begin with condition (i). Note, by definition, $W_a = \rho_a(\nu_a)$. So, we only need show that $P_a \subseteq W_a$. Fix some $s_a \in P_a$. By construction, $s_a$ is optimal under $\varpi_a$. Let $h \in H_a$ with $s_a \in S_a(h)$. If $P_b \cap S_b(h) \neq \emptyset$ then $\varpi_a(\cdot|S_b(h)) = \nu_a(\cdot|S_b(h))$ and $s_a$ is optimal under $\nu_a(\cdot|S_b(h))$ among all strategies in $S_a(h)$. (See Lemma E6.) If $P_b \cap S_b(h) = \emptyset$ then $\nu_a(\cdot|S_b(h)) = \mu_a(\cdot|S_b(h))$. So, again, $s_a$ is optimal under $\nu_a(\cdot|S_b(h))$ given all strategies in $S_a(h)$. With this, $s_a \in \rho_a(\nu_a(\cdot|\cdot))$, as required.

Next, turn to condition (ii). We need to show that $\nu_a$ strongly believes $P_b$. For this notice that if $P_b \cap S_b(h) \neq \emptyset$ then $\nu_a(P_b|S_b(h)) = \varpi_a(P_b|S_b(h)) = 1$.

Finally, we show condition (iii). Suppose $r_a$ is optimal under $\nu_a(\cdot|S_b)$. We will show that $\pi_a(r_a, s_b) = \pi_a(s_a, s_b)$ for all $(s_a, s_b) \in P_a \times P_b$. To see this, recall, $\nu_a(\cdot|S_b) = \varpi_a$. So, if $r_a$ is optimal under $\nu_a(\cdot|S_b)$ then $r_a \in X_a$. The claim now follows from the fact that $X_a \times X_b$ is constant that contains $P_a \times P_b$.

Replacing $b$ with $a$ establishes that $W_a \times W_b$ is an expansion of $P_a \times P_b$. ■

**Lemma E8** *Fix a PI game satisfying NRT. Let $(s_a, s_b)$ be a Nash equilibrium in sequentially justifiable strategies. Then there exists an EFBRS, viz. $Q_a \times Q_b$, that contains $(s_a, s_b)$.*

**Proof.** Fix a Nash equilibrium in sequentially optimal strategies, viz. $(s_a, s_b)$. Let $Q_a^0 \times Q_b^0 = \{s_a\} \times \{s_b\}$. By Lemma E4, $Q_a^0 \times Q_b^0$ satisfies the best response property. So, there is a CPS $\mu_a$ (resp. $\mu_b$) that strongly believes $\{s_b\}$ (resp. $\{s_a\}$) and $s_a$ (resp. $s_b$) is sequentially optimal under $\mu_a$ (resp. $\mu_b$). Let $Q_a^1 = \rho_a(\mu_a)$ (resp. $Q_b^1 = \rho_b(\mu_a)$). Note that $Q_a^1 \times Q_b^1$ is an expansion of $Q_a^0 \times Q_b^0$ (associated with the CPS's $\mu_a$ and $\mu_b$). Now, repeatedly apply Lemma E7 to get sets $Q_a^0 \times Q_b^0$,

$Q_a^1 \times Q_b^1$, $Q_a^2 \times Q_b^2$, ..., where each $Q_a^{m+1} \times Q_b^{m+1}$ is an expansion of $Q_a^m \times Q_b^m$. Since the game is finite, there is some $M$ with $Q_a^m \times Q_b^m = Q_a^M \times Q_b^M$ for all $m \geq M$. The set $Q_a^M \times Q_b^M$ is an EFBRS. ■

**III. Closing the Gap:** In the text, we mentioned that there is a gap between parts (i)-(ii) of Proposition 7.1. We said that we do not know if part (i) can be improved to read: If $Q_a \times Q_b$ satisfies the best response property, then each $(s_a, s_b) \in Q_a \times Q_b$ is outcome equivalent to a sequentially justifiable Nash Equilibrium. Let us better understand the problem.

Return to Lemma E1 and the proof thereof. Suppose, we strengthened the induction hypothesis, so that we can look at a sequentially justifiable Nash equilibrium of subgame 1, viz. $\left(r_a^1, r_b^1\right)$. Following the proof, we use this, to construct a Nash equilibrium $(r_a, (r_b^1, \underline{q}_b^2, \ldots, \underline{q}_b^K))$, where each $\underline{q}_b^k$ is the minimax strategy on subtree $k$. But, now we need to show that the constructed equilibrium is sequentially justifiable. Here is where the problem arises—the strategy $\underline{q}_b^k$ (on subtree $k$) may not be a best response to any strategy on that subtree. Thus, the proof breaks down. Of course, it may very well be that there is another method of proof.

In the text we mentioned a related result—namely Proposition 7.1, which speaks some to the gap. To show this result, it suffices to show the following Lemma.

**Lemma E9** *Suppose $Q_a \times Q_b$ is a constant set satisfying the best response property. Then there exists a mixed strategy Nash equilibrium, viz. $(\sigma_a, \sigma_b)$, so that:*

*(i) $Q_a \times Q_b$ is outcome equivalent to $(\sigma_a, \sigma_b)$, and*

*(ii) each $s_a \in \operatorname{Supp} \sigma_a$ (resp. $s_b \in \operatorname{Supp} \sigma_b$) is sequentially justifiable.*

**Proof.** Pick some $(r_a, r_b) \in Q_a \times Q_b$ and let $\mu_a \in \mathcal{C}(S_b)$ be a CPS so that $r_a \in \rho_a(\mu_a)$ and $\mu_a$ strongly believes $Q_b$. Set $\sigma_b = \mu_a(\cdot|S_b)$. Construct $\sigma_a$ analogously.

First, notice that $(\sigma_a, \sigma_b)$ is a mixed strategy Nash equilibrium: Begin by using the fact that $\mu_b(Q_a|S_a) = 1$ and $\mu_a(Q_b|S_b) = 1$. As such $\operatorname{Supp} \sigma_a \times \operatorname{Supp} \sigma_b \subseteq Q_a \times Q_b$. Since $Q_a \times Q_b$ is a constant set, for each $(s_a, s_b) \in \operatorname{Supp} \sigma_a \times \operatorname{Supp} \sigma_b$, $\pi(s_a, s_b) = \pi(r_a, r_b)$. So, for each $s_a \in \operatorname{Supp} \sigma_a$ and each $q_a \in S_a$,

$$\begin{aligned}
\pi_a(s_a, \sigma_b) &= \pi_a(r_a, r_b) \\
&= \pi_a(r_a, \sigma_b) \geq \pi_a(q_a, \sigma_b),
\end{aligned}$$

where the inequality holds because $r_a \in \rho_a(\mu_a)$ and $\mu_a(\cdot|S_b) = \sigma_b$. Applying an analogous argument to $b$, establishes that $(\sigma_a, \sigma_b)$ is indeed a Nash equilibrium.

Next, notice that $Q_a \times Q_b$ is outcome equivalent to $(\sigma_a, \sigma_b)$: To see this, recall that $\operatorname{Supp} \sigma_a \times \operatorname{Supp} \sigma_b \subseteq Q_a \times Q_b$ and $Q_a \times Q_b$ is a constant set. So, it is immediate that, for each $(s_a, s_b) \in Q_a \times Q_b$, $\pi(s_a, s_b) = \pi(\sigma_a, \sigma_b)$.

Lastly, notice that each $s_a \in \operatorname{Supp}\sigma_a$ is sequentially justifiable, and likewise for $b$: To see this, recall that $\operatorname{Supp}\sigma_a \times \operatorname{Supp}\sigma_b \subseteq Q_a \times Q_b$. So, if $s_a \in \operatorname{Supp}\sigma_a$, then $s_a \in Q_a$, and so $s_a$ is sequentially justifiable. ■

**Proof of Proposition 7.1.** Immediate from Lemmata E2-E9. ■

# References


[1] Aumann, R. and A. Brandenburger, “Epistemic Conditions for Nash Equilibrium,” *Econometrica*, 63, 1995, 1161-1180.

[2] Battigalli, P., “Strategic Independence and Perfect Bayesian Equilibria,” *Journal of Economic Theory*, 70, 1996, 201-234.

[3] Battigalli, P., “On Rationalizability in Extensive Form Games,” *Journal of Economic Theory*, 74, 1997, 40-61.

[4] Battigalli, P., “Rationalizability in Incomplete Information Games”, EUI working paper 17/99.

[5] Battigalli, P., and A. Prestipino, “Transparent Restrictions on Beliefs and Forward Induction Reasoning in Games with Asymmetric Information,” 2010.

[6] Battigalli, P., and M. Siniscalchi, “Interactive Beliefs, Epistemic Independence and Rationalizability,” *Research in Economics*, 53, 1999, 243-246.

[7] Battigalli, P., and M. Siniscalchi, “Hierarchies of Conditional Beliefs and Interactive Epistemology in Dynamic Games,” *Journal of Economic Theory*, 88, 1999, 188-230.

[8] Battigalli, P., and M. Siniscalchi, “Strong Belief and Forward-Induction Reasoning,” *Journal of Economic Theory*, 106, 2002, 356-391.

[9] Battigalli, P., and M. Siniscalchi, “Rationalization and Incomplete Information,” *Advances in Theoretical Economics*, 3, 2003.

[10] Battigalli, P., and M. Siniscalchi, “Interactive Epistemology in Games with Payoff Uncertainty,” *Research in Economics*, 61, 2007, 165-184.

[11] Ben Porath, E., “Rationality, Nash Equilibrium, and Backward Induction in Perfect Information Games,” *Review of Economic Studies*, 64, 1997, 23-46.

[12] Ben-Porath, E., and E. Dekel, “Signaling Future Actions and the Potential for Sacrifice, ” *Journal of Economic Theory,* 57, 1992, 36-51.

[13] Böge, W., and Th. Eisele, “On Solutions of Bayesian Games,” *International Journal of Game Theory*, 8, 1979, 193-215.

[14] Brandenburger, A., “On the Existence of a ’Complete’ Possibility Structure,” in *Cognitive Processes and Economic Behavior*, edited by M. Basili, N. Dimitri, and I. Gilboa, Routledge, 2003, 30-34.

[15] Brandenburger, A., and A. Friedenberg, “Self-Admissible Sets,” *Journal of Economic Theory*, 145, 2010, 785-811.

[16] Brandenburger, A., A. Friedenberg, and H.J. Keisler, "Admissibility in Games," *Econometrica*, 76, 2008, 307-352.

[17] Hillas, J. and E. Kohlberg, "Foundations of Strategic Equilibrium," in *Handbook of Game Theory with Economic Applications*, 2002, 1597-1663

[18] Kechris, A., *Classical Descriptive Set Theory*, Springer-Verlag, 1995.

[19] Kohlberg, E. and J.F. Mertens, "On the Strategic Stability of Equilibria," *Econometrica*, 54, 1986, 1003-1038.

[20] Kohlberg, E. and P. Reny, "Independence on Relative Probability Spaces and Consistent Assessments in Game Trees," *Journal of Economic Theory*, 75, 1997, 280 - 313.

[21] Marx, L., and J. Swinkels, "Order Independence for Iterated Weak Dominance," *Games and Economic Behavior*, 18, 1997, 219-245.

[22] Osborne, M., and A. Rubinstein, *A Course in Game Theory*, MIT Press, 1994.

[23] Pearce, D., "Rational Strategic Behavior and the Problem of Perfection," *Econometrica*, 52, 1984, 1029-1050.

[24] Purves, R., "Bimeasurable Functions," *Fundamenta Mathematicae*, 58, 1966, 149-157.

[25] Reny, P.J., "Common Belief and the Theory of Games with Perfect Information," *Journal of Economic Theory*, 59, 1993, 257-274.

[26] Renyi, A. "On a New Axiomatic Theory of Probability," *Acta Mathematica Academiae Scientiarum Hungaricae,* 6, 1955, 285-335.

[27] Shimoji, M., and J. Watson, "Conditional Dominance, Rationalizability, and Game Forms," *Journal of Economic Theory*, 83, 161-195.

[28] Srivastava, S.M., *A Course on Borel Sets*, Springer-Verlag, 1998.

[29] Stalnaker, R., "Belief Revision in Games: Forward and Backward Induction," *Mathematical Social Sciences*, 36, 1998, 31-56.

[30] Swinkels, J.M., "Independence for Conditional Probability Systems," 1994, available at www.kellogg.northwestern.edu/research/math/papers/1076.pdf.