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, contextual reasoning is 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 [6, 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 has implications for the play of the game.

Here, we study the implications of forward induction reasoning when there is a context to the game. Because there is such a context, a certain event may be "transparent" to the players. That is, the context may rule out certain beliefs. This may limit the ability of players to rationalize past behavior, and so may affect forward induction reasoning.

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

---

*We are indebted to Adam Brandenburger, John Nachbar, and Marciano Siniscalchi for many helpful conversations. We also thank Ethan Bueno de Mesquita, Alfredo Di Tillio, Alejandro Manelli, Adam Szeidl, and seminar participants at Bocconi University, Boston University, New York University, Northwestern University, and UC Berkeley for important input. Battigalli thanks MIUR and Bocconi University. Friedenberg thanks the W.P. Carey School of Business and the Olin Business School. cdfi-02-01-09.rap

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

‡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/.

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 [6, 2002] "strong belief" idea. (See also Stalnaker [26, 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 F-rationalizability concept. This solution concept is due to Battigalli-Siniscalchi [7, 2003], who referred to it as Δ-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 EFBRS's capture context-dependent forward induction reasoning. It also explains the connection to F-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 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 [16, 1986].

---

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

(See, Hillas-Kohlberg [14, 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)?

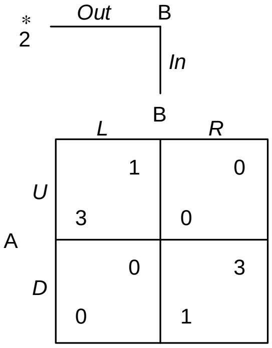


Figure 1.1

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 β^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 [6, 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 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 a different outcome than the typical 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 F-rationalizability (Battigalli-Siniscalchi [7, 2003]).

1.4 Rationality and Common Strong Belief of Rationality

Let us begin with an arbitrary epistemic game. Refer to Figure 1.2. Here, **R0SBR** is the set of states at which each player is rational. **R1SBR** is the set of states at which each player is rational and strongly believes "the other player is rational." More generally, **RmSBR** is the set of states at which there is **rationality and m^{th}-order strong belief of rationality** (**RmSBR**).

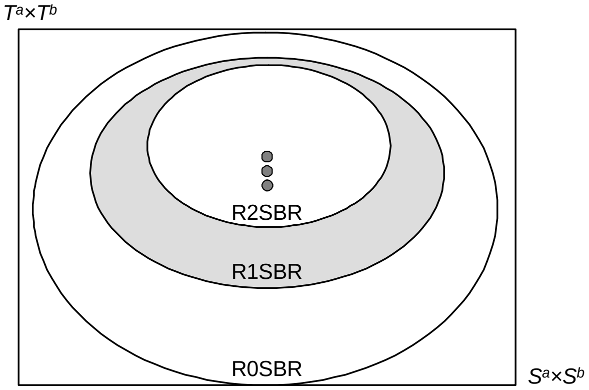


Figure 1.2

We are interested in the set of states consistent with **rationality and common strong belief of rationality** (**RCSBR**). This is the intersection of the sets RmSBR, 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 consistent with RmSBR, for each m. We will use this to identify two facts about s^a.

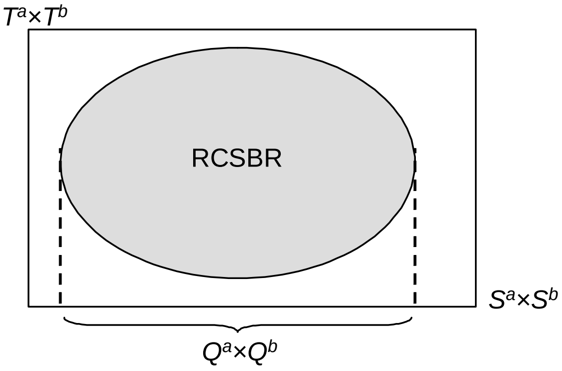


Figure 1.3

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 R0SBR event for Bob, the R1SBR event for Bob, the R2SBR event for Bob, etc. So, by a conjunction property of strong belief, t^a strongly believes the RCSBR event for Bob. 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.

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

$Q^a \times Q^b$.

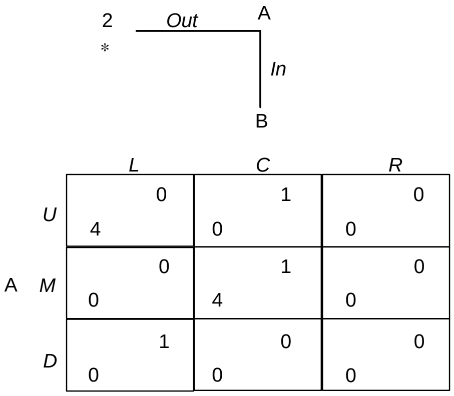


Figure 1.4

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.

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, $(In\text{-}Up, t^a)$ and $(In\text{-}Middle, t^a)$ are also rational. Indeed, since t^a strongly believes each of the RmSBR events for Bob, both $(In\text{-}Up, t^a)$ and $(In\text{-}Middle, t^a)$ must be consistent with RCSBR.

Next, consider an RCSBR strategy-type pair for Bob, viz. $\left(s^b, t^b\right)$. 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 $\left(s^b, t^b\right)$ 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 $\mu^b\left(s^b\right)$ (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 β^a and β^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 [6, 2002] for a formal statement.

1.7 F-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

---

[2] The condition that β^a and β^b are onto is known as completeness. It is due to Brandenburger [11, 2003].

gives a restriction on Bob's first-order beliefs, i.e., his beliefs about what Ann will choose. Let F^b represent this restriction on first-order beliefs. So, F^b is a subset of the CPS's on S^a and, in our example, F^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 F^a for the set of all CPS's on S^b.

With $F = F^a \times F^b$ 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-Left* and *In-Right* for Bob, since these strategies are not sequentially optimal under the CPS in F^b. We do not eliminate any of Ann's strategies, since they are each sequentially optimal under some CPS (in F^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 F^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 F^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 **F-rationalizability**. It is due to Battigalli-Siniscalchi [7, 2003], who referred to it as Δ-rationalizability.[3] The procedure begins by fixing a set of **first-order beliefs**, i.e., a set $F = F^a \times F^b$, where F^a is a set of CPS's on S^b and F^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 F^a (resp. F^b). On round two, it further eliminates any strategy of Ann (resp. Bob) that is not sequentially optimal under a CPS in F^a (resp. F^b) that strongly believes the round-one strategies of Bob (resp. Ann). And so on.

Note, there may be many F-rationalizable sets—each of which is obtained by beginning the procedure with a different set of first-order beliefs $F = F^a \times F^b$. Collecting all the F-rationalizable sets together, we get the solution concept of F-rationalizability.

1.8 An Alternate Characterization Theorem

In Section 1.7, we considered a particular set of first-order beliefs, and we computed the associated F-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. $F = F^a \times F^b$, we can always find an epistemic structure so that the F-rationalizable strategy set is the set of strategies consistent with RCSBR. In particular, we will see that the F-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-Sinsicalchi [7, 2003] introduced the concept to study a different problem from the one studied here—specifically, in their problem, the set F is given to the analyst. See Section 8a.

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

Indeed, beginning with an epistemic structure and the RCSBR strategy set, we will get some F-rationalizable set. But, importantly, the approach taken to find this F-rationalizable set may be different from the approach we took in Section 1.7. To understand why, let us mimic the route we took there.

In Section 1.7, we begin with an epistemic structure and use the structure itself to form the set $F^a \times F^b$. 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 F^a. Construct the set F^b analogously. Note, here, the strategies that survive one round of F-rationalizability are exactly the strategies that are consistent with R0SBR. 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 F-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 F-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 F-rationalizability. So $\beta^a(t^a)$ need not strongly believe this latter event. As such, we loose any relationship between the RCSBR strategies and the $F^a \times F^b$-rationalizable strategy set.

But, there is another route, that instead uses the EFBRS properties to form the set $F^a \times F^b$. 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 F^a to be the set of such CPS's, i.e., one for each $s^a \in Q^a$, and construct F^b similarly. Now we do have an equivalence between the RCSBR strategies and the F-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 F-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 $F = F^a \times F^b$ so that the set of strategies consistent with RCSBR is exactly the F-rationalizable strategy set.

(ii) Fix an extensive-form game and a set of first-order beliefs $F = F^a \times F^b$. Then there exists an epistemic type structure, so that the set of strategies consistent with RCSBR

is exactly the F-rationalizable strategy set.

Indeed, the solution concept of F-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 F-rationalizable set is associated with a particular type structure. To see this, fix a type structure and let $G^a \times G^b$ be the set of all first-order beliefs associated with that particular structure. The RCSBR strategies form some $\left(F^a \times F^b\right)$-rationalizable set, but this set may be distinct from the $\left(G^a \times G^b\right)$-rationalizable set.

While, in general, it may not be obvious which F-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, as in the "lady's choice convention" discussed above. Battigalli-Prestipino [4, 2009] provide a formal statement. They show that, in this case, the RCSBR strategy set does correspond to the $\left(G^a \times G^b\right)$-rationalizable strategy set. (See Section 8a below.)

1.9 The EFBRS Properties

We have seen that the F-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. $F = F^a \times F^b$, so that the F-rationalizable set is $Q^a \times Q^b$.

While the EFBRS and F-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, and, there, the EFBRS properties will play an important role. We will also see that the properties are analogous to the properties of a self-admissible set (Brandenburger-Friedenberg-Keisler [13, 2008]), and so there are some interesting connections in terms of applications.

In Section 8c, we return to further discuss the EFBRS vs. F-rationalizability definitions.

2 The Game

We consider finite extensive form games of perfect recall. We write Γ for such a game. The definition we consider is similar to that in Osborne-Rubinstein [19, 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

---

[4] We choose this definition to allow simultaneous moves—in particular, this definition incorporates repeated games as a special case. 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.

choices for Ann and Bob. More formally, a **history** is either (i) the empty sequence, written ϕ, or (ii) a sequence of choice pairs $(c_1, \ldots, c_K)$, where each $c_k = (c_k^a, c_k^b) \in C^a \times C^b$. Note, histories have the property that, if $(c_1, \ldots, c_K)$ is a history then so is $(c_1, \ldots, c_L)$, 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** for 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 non-terminal histories, with 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 family of **information sets** for a (resp. b). (Note, perfect recall imposes further requirements on H^a and H^b. See Osborne-Rubinstein [19, 1994; Definition 203.3].)

Write Z for the set of terminal histories for 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 payoff vector as an **outcome** of the game. Two strategy profiles, (s^a, s^b) and (r^a, r^b), are **outcome equivalent** if $\pi(s^a, s^b) = \pi(r^a, r^b)$.

For each information set 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 [6, 2002].

Throughout, let Ω be a separable metrizable space and let $\mathcal{B}(\Omega)$ the Borel σ-algebra on Ω. 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 Ω, and endow $\mathcal{P}(\Omega)$ with the topology of weak convergence.

Definition 3.1 (Renyi [23, 1955]) *Fix a separable metrizable space Ω 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 $\emptyset \neq \mathcal{E} \subseteq \mathcal{B}(\Omega)$ a **collection of conditioning events** for Ω.*

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 $\nu(\cdot|\cdot)$ is a conditional probability system on $(\Omega_1, \mathcal{E})$. When $\nu(\cdot|\cdot)$ is defined in this way, write $\nu(\cdot|\cdot) = \text{marg}_{\Omega_1} \mu(\cdot|\cdot)$. No confusion should result.

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

$$\left\langle S^a, S^b; \mathcal{S}^a, \mathcal{S}^b; T^a, T^b; \beta^a, \beta^b \right\rangle,$$

*where T^a (resp. T^b) is a nonempty separable metrizable space and $\beta^a : T^a \to \mathcal{C}\left(S^b \times T^b\right)$ (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 [5, 1999] canonical construction of a type structure that contains all hierarchies of conditional beliefs. Lets us look at the so-called self-evident events within this structure. Loosely, these are events $E \in \mathcal{B}\left(S^a \times T^a \times S^b \times T^b\right)$, 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., $\pi^a(s^a,\mu^a) = \sum_{s^b \in S^b} \pi^a(s^a,s^b)\,\mu^a(s^b)$. (Notice, the measure μ^a on S^b reflects a belief by a about b, so we write $\mu^a \in \mathcal{P}(S^b)$.)

Definition 4.1 *Fix $X^a \subseteq S^a$ and $s^a \in X^a$. Say s^a is **optimal under** $\mu^a \in \mathcal{P}(S^b)$ **given** X^a if $\pi^a(s^a,\mu^a) \geq \pi^a(r^a,\mu^a)$ 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 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 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 $\operatorname{marg}_{S^b} \beta^a(t^a)(\cdot|\cdot)$.*

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 [6, 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 μ **strongly believes** E if, for each $F \in \mathcal{E}$, $E \cap F \neq \emptyset$ implies $\mu(E|F) = 1$.*

We point out two general properties about strong belief.

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

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

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

$$\mathrm{SB}^a(E^b) = S^a \times \{t^a \in T^a : t^a \text{ strongly believes } E^b\},$$

and $\mathrm{CSB}^a\left(E^b\right) = E^b \cap \mathrm{SB}^a\left(E^b\right)$. That is, $\mathrm{SB}^a\left(E^b\right)$ is the event "Ann strongly believes E^b" and $\mathrm{CSB}^a\left(E^b\right)$ is the event "Ann strongly believes E^b and E^b is in fact correct." Given a product set $E \in \mathcal{B}\left(S^a \times T^a \times S^b \times T^b\right)$, viz. $E = E^a \times E^b$, write $\mathrm{SB}(E) = \mathrm{SB}^a\left(E^b\right) \times \mathrm{SB}^b(E^a)$ and $\mathrm{CSB}(E) = \mathrm{CSB}^a\left(E^b\right) \times \mathrm{CSB}^b(E^a)$.

Note, $\mathrm{SB}(\cdot) = \mathrm{SB}^a(\cdot) \times \mathrm{SB}^b(\cdot)$ can be viewed as a mutual strong belief operator. Then, $\mathrm{CSB}(\cdot) = \mathrm{CSB}^a(\cdot) \times \mathrm{CSB}^b(\cdot)$ 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 $E \in \mathcal{B}\left(S^a \times T^a \times S^b \times T^b\right)$, write $\mathrm{CSB}^0(E) = E$ and, for each $m \geq 0$, define $\mathrm{CSB}^{m+1}(E) = \mathrm{CSB}\left(\mathrm{CSB}^m(E)\right)$.[6] So,

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

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

and so on. Note that

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

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* $(s^a, t^a, s^b, t^b) \in \bigcap_m \mathrm{CSB}^m\left(R^a \times R^b\right)$.

5 Characterization Theorems

We now turn to characterizing RCSBR. For this it will be useful to introduce a **best reply correspondence**, viz. $\rho^a : \mathcal{C}\left(S^b \times T^b, \mathcal{S}^b\right) \to 2^{S^a}$, where $\rho^a(\mu^a(\cdot|\cdot))$ is the set of strategies that are sequentially optimal under $\mu^a(\cdot|\cdot)$. 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* $\mu^a(\cdot|\cdot) \in \mathcal{C}\left(S^b, \mathcal{S}^b\right)$ *so that:*

(i) $s^a \in \rho^a(\mu^a(\cdot|\cdot))$,

(ii) $\mu^a(\cdot|\cdot)$ *strongly believes* Q^b, *and*

(iii) $\rho^a(\mu^a(\cdot|\cdot)) \subseteq Q^a$.

And similarly with a and b reversed.

Theorem 5.1 *Fix an extensive-form game* Γ.

(i) For any Γ*-based type structure,* $\mathrm{proj}_{S^a \times S^b} \bigcap_m \mathrm{CSB}^m\left(R^a \times R^b\right)$ *is an EFBRS.*

---

[6] Note, we use superscripts both for players and levels of reasoning. No confusion should result.

(ii) Fix an EFBRS $Q^a \times Q^b$. There exists a Γ-based type structure, so that $Q^a \times Q^b = \mathrm{proj}_{S^a \times S^b} \bigcap_m \mathrm{CSB}^m \left(R^a \times R^b\right)$.

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 $R^{a,(m+1)} = R^{a,m} \cap \mathrm{SB}^a \left(R^{b,m}\right)$ (resp. $R^{b,(m+1)} = R^{b,m} \cap \mathrm{SB}^b \left(R^{a,m}\right)$). Then, a standard induction argument gives that $\mathrm{CSB}^m \left(R^a \times R^b\right) = R^{a,m} \times R^{b,m}$, for each m. It follows that $\bigcap_m \mathrm{CSB}^m \left(R^a \times R^b\right) = \bigcap_m \left(R^{a,m} \times R^{b,m}\right)$. We make use of this below.

Proof. Begin with part (i). Fix a Γ-based type structure. If $\bigcap_m \mathrm{CSB}^m \left(R^a \times R^b\right) = \emptyset$ then the result is immediate. So, suppose $\bigcap_m \mathrm{CSB}^m \left(R^a \times R^b\right) \neq \emptyset$.

Fix $(s^a, s^b) \in \mathrm{proj}_{S^a \times S^b} \bigcap_m \mathrm{CSB}^m \left(R^a \times R^b\right)$. Then there exists (t^a, t^b) such that

$$\left(s^a, t^a, s^b, t^b\right) \in \bigcap_m \mathrm{CSB}^m \left(R^a \times R^b\right) = \bigcap_m (R^{a,m} \times R^{b,m}).$$

We will show that the CPS $\mathrm{marg}_{S^b} \beta^a \left(t^a\right)$ satisfies conditions (i)-(iii) for s^a. A similar argument holds for s^b.

First note,

$$(s^a, t^a) \in \rho^a \left(\mathrm{marg}_{S^b} \beta^a \left(t^a\right)\right) \times \{t^a\} \subseteq R^a.$$

Now use the fact that t^a strongly believes each R^b_m to get that

$$\rho^a \left(\mathrm{marg}_{S^b} \beta^a \left(t^a\right)\right) \times \{t^a\} \subseteq \bigcap_m R^{a,m}.$$

So, $s^a \in \rho^a \left(\mathrm{marg}_{S^b} \beta^a \left(t^a\right)\right) \subseteq \mathrm{proj}_{S^a} \bigcap_m R^{a,m}$, establishing conditions (i) and (iii). Next note that, using the Conjunction Property of strong belief (Property 4.1), $\beta^a \left(t^a\right)$ strongly believes $\bigcap_m R^{b,m}$. Using the Marginalization Property (Property 4.2), $\mathrm{marg}_{S^a} \beta^a \left(t^a\right)$ strongly believes $\mathrm{proj}_{S^b} \bigcap_m R^{b,m}$. This establishes condition (ii).

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 $\mu^a \left(t^a\right)(\cdot|\cdot) \in \mathcal{C}\left(S^b\right)$ satisfying conditions (i)-(iii) of an EFBRS. Now construct a CPS $\beta^a \left(t^a\right) \in \mathcal{C}\left(S^b \times T^b, \mathcal{S}^b \otimes T^b\right)$ as follows. If $Q^b \cap S^b (h) \neq \emptyset$, set $\beta^a \left(t^a\right) \left(\left(t^b, t^b\right) | S^b (h) \times T^b\right) = \mu^a \left(t^a\right) \left(t^b | S^b (h)\right)$ for each $t^b \in Q^b = T^b$. Next, fix some arbitrary element $t^b_* \in T^b$. Then, if $Q^b \cap S^b (h) = \emptyset$, set $\beta^a \left(t^a\right) \left(\left(s^b, t^b_*\right) | S^b (h) \times T^b\right) = \mu^a \left(t^a\right) \left(s^b | S^b (h)\right)$ 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 $\beta^a \left(t^a\right)$ 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,

$$\begin{aligned}
\beta^a(t^a)\left(E^b|S^b(i)\times T^b\right) &= \mu^a(t^a)\left(\left\{t^b \in Q^b : \left(t^b, t^b\right)\in E^b\right\}|S^b(i)\right)\\
&= \mu^a(t^a)\left(\left\{t^b \in Q^b : \left(t^b, t^b\right)\in E^b\right\}|S^b(h)\right)\times \mu^a(t^a)\left(S^b(h)|S^b(i)\right)\\
&= \mu^a(t^a)\left(\left\{t^b \in Q^b : \left(t^b, t^b\right)\in E^b\right\}|S^b(h)\right)\times \mu^a(t^a)\left(Q^b\cap S^b(h)|S^b(i)\right)\\
&= \beta^a(t^a)\left(E^b|S^b(h)\times T^b\right)\times \beta^a(t^a)\left(S^b(h)\times T^b|S^b(i)\times T^b\right),
\end{aligned}$$

where the first and fourth lines follow from the construction, the second follows from the fact that $\mu^a(t^a)(\cdot|\cdot)$ is a CPS, and the third line follows from the fact that $\mu^a(t^a)\left(Q^b|S^b(h)\right) = 1$ (i.e., $\mu^a(t^a)(\cdot|\cdot)$ 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 $\mu^a(t^a)\left(\operatorname{proj}_{S^b} E^b|S^b(i)\right) = 0$ and $\mu^a(t^a)\left(S^b(h)|S^b(i)\right) = 0$. (This uses the fact that $\mu^a(t^a)\left(Q^b|S^b(i)\right) = 1$.) So, here too,

$$\begin{aligned}
\beta^a(t^a)\left(E^b|S^b(i)\times T^b\right) &= \beta^a(t^a)\left(E^b|S^b(h)\times T^b\right)\times \beta^a(t^a)\left(S^b(h)\times T^b|S^b(i)\times T^b\right)\\
&= 0.
\end{aligned}$$

Finally, suppose $Q^b \cap S^b(i) = \emptyset$. Here,

$$\begin{aligned}
\beta^a(t^a)\left(E^b|S^b(i)\times T^b\right) &= \mu^a(t^a)\left(\left\{s^b : \left(s^b, t^b_*\right)\in E^b\right\}|S^b(i)\right)\\
&= \mu^a(t^a)\left(\left\{s^b : \left(s^b, t^b_*\right)\in E^b\right\}|S^b(h)\right)\times \mu^a(t^a)\left(S^b(h)|S^b(i)\right)\\
&= \beta^a(t^a)\left(E^b|S^b(h)\times T^b\right)\times \beta^a(t^a)\left(S^b(h)\times \{t^b_*\}|S^b(i)\times T^b\right)\\
&= \beta^a(t^a)\left(E^b|S^b(h)\times T^b\right)\times \beta^a(t^a)\left(S^b(h)\times T^b|S^b(i)\times T^b\right),
\end{aligned}$$

as required.

We will conclude the proof by showing

$$Q^a = \bigcup_{t^a\in T^a}\left[\rho^a\left(\operatorname{marg}_{S^b}\beta^a(t^a)\right)\right] \tag{5.1}$$

$$R^{a,m} = \bigcup_{t^a\in T^a}\left[\rho^a\left(\operatorname{marg}_{S^b}\beta^a(t^a)\right)\times\{t^a\}\right] \quad \text{for each } m, \tag{5.2}$$

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

To show Equation 5.1: Recall, for each $t^a \in T^a = Q^a$, $\mu^a(t^a)(\cdot|\cdot) = \operatorname{marg}_{S^b}\beta^a(t^a)(\cdot|\cdot)$. So, it is immediate from the construction that $Q^a \subseteq \bigcup_{t^a\in T^a}\rho^a\left(\operatorname{marg}_{S^b}\beta^a(t^a)\right)$. Conversely, fix any strategy s^a in $\bigcup_{t^a\in T^a}\rho^a\left(\operatorname{marg}_{S^b}\beta^a(t^a)\right)$. 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

$R^{b,m} \cap \left[S^b(h) \times T^b\right] \neq \emptyset$. Observe that

$$\begin{aligned} \left[\operatorname{proj}_{S^b} R^{b,m}\right] \cap S^b(h) &= \left[\bigcup_{t^b \in T^b} \rho^b\left(\operatorname{marg}_{S^a} \beta^b\left(t^b\right)\right)\right] \cap S^b(h) \\ &= Q^b \cap S^b(h). \end{aligned}$$

(The first equality follows from the induction hypothesis for b. The second equality follows from Equation 5.1.) Since $R^{b,m} \cap \left[S^b(h) \times T^b\right] \neq \emptyset$, it follows that $Q^b \cap S^b(h) \neq \emptyset$, and so $\mu^a(t^a)\left(Q^b | S^b(h)\right) = 1$. (Here, we use part (ii) of the definition of an EFBRS.) So, by construction, $\beta^a(t^a)\left(R^{b,m} | S^b(h) \times T^b\right) = 1$, as required. ■

Now, we turn to F-rationalizability. Let F^a (resp. F^b) be a non-empty subset of $\mathcal{C}\left(S^b\right)$ (resp. $\mathcal{C}(S^a)$), i.e. a set of first-order beliefs of Ann (resp. Bob). Call $F = F^a \times F^b$ a **set of first-order beliefs**. Set $S_F^{a,0} = S^a$ and $S_F^{b,0} = S^b$. Inductively define $S_F^{a,m}$ and $S_F^{b,m}$ as follows: Let $S_F^{a,m+1}$ be the set of all $s^a \in S_F^{a,m}$ so that, there is some CPS $\mu^a \in F^a$ with (i) $s^a \in \rho^a(\mu^a)$ and (ii) μ^a strongly believes $S_F^{b,m}$. And, likewise, with a and b interchanged.

Definition 5.2 (Battigalli-Siniscalchi [7, 2003]) *Call $S_F^a = \bigcap_{m \geq 0} S_F^{a,m}$ (resp. $S_F^b = \bigcap_{m \geq 0} S_F^{b,m}$) the* ***F-rationalizable strategies of Ann*** *(resp.* ***Bob****). Call $S_F^a \times S_F^b$ the* ***F-rationalizable strategy set****.*

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

Proposition 5.1 *Fix an extensive-form game Γ.*

(i) Given a set of first-order beliefs, viz. $F = F^a \times F^b$, $S_F^a \times S_F^b$ is an EFBRS.

(ii) Given an EFBRS, viz. $Q^a \times Q^b$, there exists a set of first-order beliefs, viz. $F = F^a \times F^b$, so that $S_F^a \times S_F^b = Q^a \times Q^b$.

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

Corollary 5.1 *Fix an extensive-form game Γ.*

(i) For any Γ-based type structure, there exists a set of first-order beliefs, viz. $F = F^a \times F^b$, so that $S_F^a \times S_F^b = \operatorname{proj}_{S^a \times S^b} \bigcap_m \mathrm{CSB}^m\left(R^a \times R^b\right)$.

(ii) Fix a set of first-order beliefs, viz. $F^a \times F^b$. Then there exists a Γ-based structure, so that $S_F^a \times S_F^b = \operatorname{proj}_{S^a \times S^b} \bigcap_m \mathrm{CSB}^m\left(R^a \times R^b\right)$.

Now for the proof.

Proof of Proposition 5.1. Begin with part (i), and some set of first-order beliefs, viz. $F = F^a \times F^b$. Note, there exists some M with $S_F^a \times S_F^b = S_F^{a,M} \times S_F^{b,M}$. Fix $s^a \in S_F^a = S_F^{a,M+1}$. By

Lemma C1 in Appendix C, we can find a CPS $\mu^a(\cdot|\cdot)$ so that $s^a \in \rho^a(\mu^a(\cdot|\cdot))$ and $\mu^a(\cdot|\cdot)$ strongly believes each $S_F^{b,m}$ for $m \leq M$. Thus, s^a satisfies conditions (i)-(ii) of Definition 5.1. Moreover, if $r^a \in \rho^a(\mu^a(\cdot|\cdot))$, then r^a is optimal under a CPS that strongly believes each $S_F^{b,m}$, for $m \leq M$. As such, $r^a \in S_F^{a,m}$ for each $m \leq M$, establishing that $r^a \in S_F^a$.

Now turn to part (ii). Fix $Q^a \times Q^b$. For each $s^a \in Q^a$, there exists some CPS $\mu^a(s^a) \in \mathcal{C}(S^b)$ satisfying conditions (i)-(iii) of an EFBRS. Take F^a so that, for each $s^a \in Q^a$, F^a contains exactly one such CPS $\mu^a(s^a)$ (and no other CPS's). And, likewise, with a and b interchanged. We will show that, for each $m \geq 1$, $S_F^{a,m} \times S_F^{b,m} = Q^a \times Q^b$. This will establish the result.

The proof is by induction. Begin with $m = 1$. Certainly $Q^a \subseteq S_F^{a,1}$. Fix $s^a \in S_F^{a,1}$. Then there exists some $\mu^a(r^a) \in S_F^{a,1}$ so that s^a is sequentially optimal under $\mu^a(r^a)$. This CPS $\mu^a(r^a)$ is associated with some $r^a \in Q^a$, i.e., so that r^a and $\mu^a(r^a)$ jointly satisfy conditions (i)-(iii) of an EFBRS. So, we can apply condition (iii) of an EFBRS to get that $s^a \in Q^a$. And, likewise, for b.

Now assume the results holds for $m \geq 2$. We will show it also holds for $m+1$. Fix $s^a \in Q^a = S_F^{a,m}$. Then, using conditions (i)-(ii) of an EFBRS, there exists some $\mu^a(s^a)$ so that $s^a \in \rho^a(\mu^a(s^a))$ and $\mu^a(s^a)$ strongly believes $Q^b = S_F^{b,m}$. So, certainly, $Q^a \subseteq S_F^{a,(m+1)}$. Conversely, fix some $s^a \in S_F^{a,(m+1)}$. Then, there exists a CPS $\mu^a(r^a) \in F^a$ so that $s^a \in \rho^a(\mu(r^a))$ and $\mu^a(r^a)$ strongly believes $S_F^{b,m}$. Again, since $\mu^a(r^a)$ satisfies conditions (i)-(iii) of an EFBRS for some $r^a \in Q^a$, it follows that $\rho^a(\mu^a(r^a)) \subseteq Q^a$, and so $s^a \in Q^a$. ■

Let us comment on the proof. Begin with some finite set of first-order beliefs, viz. F. Proposition 5.1(i) says that $S_F^a \times F_F^b$ is an EFBRS. Conversely, begin with some EFBRS. The proof of Proposition 5.1(ii) says that we can find a finite set of first-order beliefs, viz. F, so that $S_F^a \times S_F^b$ is this EFBRS. With this in mind:

Remark 5.1 *The set of all F-rationalizable strategy sets is the set*

$$\{S_F^a \times S_F^b : F = F^a \times F^b \subseteq \mathcal{C}(S^b) \times \mathcal{C}(S^a) \text{ is finite}\}.$$

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

6 Examples

We have seen that RCSBR is characterized by the EFBRS concept or, equivalently, by the F-rationalizability concept. Now, we ask what this gives in games of interest. For this, it will be helpful to make use of the EFBRS properties. (In fact, we will only need to make use of Properties (i)-(ii).) Let us begin with Centipede.

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

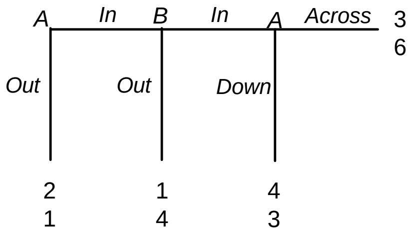


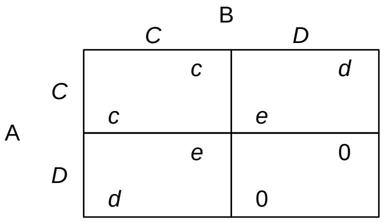
Figure 6.1

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.*



		B	
		C	D
A	C	c c	d e
	D	e d	0 0

d > c > 0 > e

Figure 6.2

Let $Q^a \times Q^b$ *be a nonempty EFBRS. Then each* $\left(s^a, s^b\right) \in Q^a \times Q^b$ *results in the Defect-Defect path.*

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* $\mu^a(s^a)\left(Q^b|S^b(h)\right) = 1$*. 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,* $s^a(h) = D$*. And, likewise, with* a *and* b *reversed.*

Now turn to the first period, and suppose, contra hypothesis, $s^a(\phi) = C$ *for some* $s^a \in Q^a$*. Note, for each* $s^a \in Q^b$*,* $\left(s^a, s^b\right)$ *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* $\mu^a(s^a)\left(\cdot|S^b\right)$*.*

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

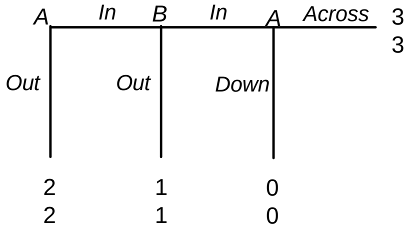


Figure 6.3

Example 6.3 *Consider the coordination game in Figure 6.3. Here, there are three EFBRS's, namely* $\{Out\} \times \{Out\}$*,* $\{Out\} \times \{Out, In\}$*, and* $\{In\text{-}Across\} \times \{In\}$*. The EFBRS* $\{In\text{-}Across\} \times \{In\}$ *corresponds to the backward induction outcome, but the other two EFBRS's do not.*

Taken together with the Main Theorem (Theorem 5.1), we get 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 maintain a hypothesis that Ann is irrational. In this case, Bob may very well maintain a hypothesis 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

Here, we turn to analyzing EFBRS's in perfect-information games, i.e., games where each information set is a singleton and there is at most one active player at each information set. We've seen two examples of perfect-information games, namely Examples 6.1 and 6.3. In the former case, each EFBRS was outcome equivalent to the backward induction outcome. Of course, for that game, the Nash and backward induction outcomes coincide. On the other hand, in Example 6.3, one EFBRS corresponds to the backward induction outcome, but others do not. However, there we do get that the EFBRS's correspond (exactly) to the 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. Let us begin with two such conditions.

Definition 7.1 (Brandenburger-Friedenberg [12, 2004]) *A game satisfies the* ***single payoff condition (SPC)*** *if, for all terminal nodes z and z', the following holds: If a moves at the last common predecessor of z and z', then $\Pi^a(z) = \Pi^a(z')$ implies $\Pi^b(z) = \Pi^b(z')$. And similarly with a and b interchanged.*

Definition 7.2 (Battigalli [3, 1997]) *A game satisfies* ***no relevant ties (NRT)*** *if, for all terminal nodes z and z', the following holds: If a moves at the last common predecessor of z and z', then $\Pi^a(z) \neq \Pi^a(z')$. And similarly for b.*

A game with no ties satisfies NRT, but the converse does not hold. Reny's [22, 1993; Figure 1] Take-It-Or-Leave-It game is one such example. Likewise, a game satisfying NRT also satisfies SPC. Yet, many games of interest satisfy SPC, but fail NRT. Zero sum games satisfy SPC, but need not satisfy NRT. In perfect-information games, SPC is equivalent to "transference of decision-maker indifference" (Marx-Swinkels [18, 1997]).[7]

Now let us state the connection:

Proposition 7.1

(i) Fix a PI game Γ satisfying SPC. If $Q^a \times Q^b$ is an EFBRS then, there exists a 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 Γ satisfying NRT. If (s^a, s^b) is a 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 Γ satisfying SPC, and an epistemic type structure. Each state at which there is RCSBR is outcome equivalent to some Nash equilibrium.

---

[7] 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 Γ *satisfying NRT, and a Nash equilibrium in sequentially justifiable strategies. Then, there exists an epistemic structure and a state thereof at which there is RCSBR and the Nash equilibrium is played.*

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 [12, 2004], though now making use of the EFBRS properties. (The proof in [12, 2004] makes use of properties of self-admissible sets. See 8c below.) Indeed, we only need use properties (i)-(ii) of Definition 5.1.

The converse, i.e., part (ii), is novel. 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 *Consider the game in Figure 7.1, which shows that part (i) of Proposition 7.1 is false, absent the SPC condition.*[8]

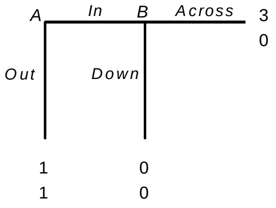


Figure 7.1

Here, $\{In\} \times \{Across, Down\}$ *is an EFBRS, but* $(In, Down)$ *is not outcome equivalent to a Nash Equilibrium.*

---

[8] This example was suggested by Drew Fudenberg. It is also Example 6.2 in Brandenburger-Friedenberg [12, 2004].

In the above example, 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 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 7.2, which satisfies SPC.*

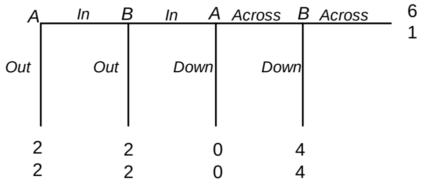


Figure 7.2

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* μ^b *strongly believes* Q^a*, then* μ^b *must assign probability one to In-Across conditional on the event* $\{In\text{-}Across, In\text{-}Down\}$*. 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, let us note a gap between parts (i)-(ii) of Proposition 7.1. In particular, part (i) says that starting from an EFBRS we can get a Nash outcome, while part (ii) says that starting from a sequentially justifiable Nash equilibrium, we can get an EFBRS.

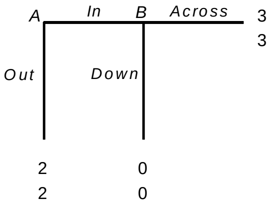


Figure 7.3

We cannot improve part (ii) to say that, starting from a Nash outcome, we get an EFBRS. To see this, refer to Figure 7.3. 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 Nash equilibrium in sequentially justifiable strategies? We do not know. In Appendix E, we elaborate on this issue.

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 F-rationalizability solution concept characterizes RCSBR across all type structures.

Note, carefully, that Battigalli-Siniscalchi [7, 2003] introduced the F-rationalizability procedure 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 restriction on players' first-order beliefs.[9] Specifically, the analyst is given a set of first-order beliefs, viz. $F^a \times F^b$, which satisfies two conditions: (i) each type t^a has $\text{marg}_{S^b}\, \beta^a(t^a)$ contained in F^a and (ii) for each CPS μ^a on $S^b \times T^b$ with $\text{marg}_{S^b}\, \mu^a \in F^a$, there is a type t^a with $\beta^a(t^a) = \mu^a$. And, likewise, with a and b interchanged. Battigalli-Siniscalchi [8, 2007] and Battigalli-Prestipino [4, 2009] provide (distinct) formal treatments along these lines. They each get the $F^a \times F^b$-rationalizable strategy set, as an output.

b. Two Characterization Theorems: We have provided two characterizations of RCSBR—namely, the EFBRS solution concept and the F-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 F-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 F-rationalizability procedure is useful. (See part a of this Section.)

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. Moreover, there is a sense in which it may be "easier" to compute the solution concept, when beginning from the EFBRS definition vs. the F-rationalizability definition. In particular, to compute the concept according to the F-rationalizability definition, we must begin with each finite set of first-order beliefs and run the F-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 $Q^a \times Q^b$.[10]

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 [12, 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

---

[9] This case is perhaps more relevant for applications of the theory of games with incomplete information, which is the focus of Battigalli-Siniscalchi [7, 2003]: 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.

[10] 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.

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.

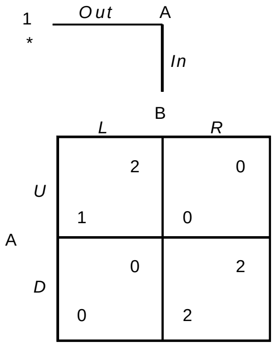


Figure 8.1

Putting these considerations together, we can have an EFBRS that is not an SAS, and an SAS 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 $\{In{-}Down\} \times \{Right\}$. (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 $\{In{-}Middle\} \times \{Center\}$. (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 [20, 1984]) $Q^a \times Q^b$. The definition of a best response set can be given both in terms of rationalizability (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 [13, 2008]) concept, we can also provide a definition both in terms of rationalizability and in terms of dominance.

Here, we have provided a rationalizability 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 [24, 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 μ^a that—in addition to satisfying conditions (i)-(ii)—also satisfies the requirement that, if r^a is sequentially optimal under μ^a, then $r^a \in Q^a$. Of course, strategies r^a that are sequentially optimal under μ^a are conditionally undominated (see [24, 1998; Lemma 2]), but a conditionally undominated strategy need not be sequentially optimal under the given CPS μ^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 [13, 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. In this case, 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 $F^a \times F^b$ to be the set of all CPS's.) As such, there exists a non-empty EFBRS. See Battigalli [3, 1998; 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. Such property is easy to state in games with observable deviators. (See, e.g., Battigalli [2, 1996].) Battigalli [2, 1996], Kohlberg-Reny [17, 1997], Stalnaker [26, 1998], and Swinkels [27, 1994] each address this issue, 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.

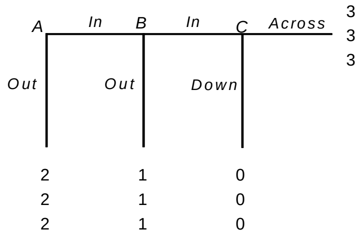


Figure 8.2

How does this affect our analysis of games? Amend Figure 6.3, to 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.[11]

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 [12, 2004] SAS analysis. But there is another important connection to be made, namely to Ben Porath [9, 1997; Theorem 2].

The starting point in Ben Porath [9, 1997] is "rationality and common initial belief of rationality." (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.) This 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

---

[11] See Stalnaker [26, 1998] for a related idea.

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* μ ***believes*** E *if, for each* $F \in \mathcal{E}$, $\mu(E|F) = 1$.

In what follows, fix a game Γ and a Γ-based type structure $\mathcal{T} = \left\langle S^a, S^b; \mathcal{S}^a, \mathcal{S}^b; T^a, T^b; \beta^a, \beta^b \right\rangle$.

Definition A2 *Say a type* $t^a \in T^a$ ***believes*** $E^b \in \mathcal{B}\left(S^b \times T^b\right)$ *if* $\beta^a(t^a)$ *believes* E^b.

Given an event $E^b \in \mathcal{B}\left(S^b \times T^b\right)$, write $\mathrm{B}^a\left(E^b\right)$ for $S^a \times \{t^a \in T^a : t^a \text{ believes } E^b\}$. When $E^a \times E^b \in \mathcal{B}\left(S^a \times T^a \times S^b \times T^b\right)$, write $\mathrm{B}\left(E^a \times E^b\right) = \mathrm{B}^a\left(E^b\right) \times \mathrm{B}^b\left(E^a\right)$. 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 Appendix A.1 (Monotonicity) *Fix events* $E^a \times E^b$ *and* $F^a \times F^b$ *in* $\mathcal{B}\left(S^a \times T^a \times S^b \times T^b\right)$. *If* $E^a \times E^b \subseteq F^a \times F^b$, *then* $\mathrm{B}\left(E^a \times E^b\right) \subseteq \mathrm{B}\left(F^a \times F^b\right)$.

Property Appendix A.2 (Conjunction) *Fix a sequence of events* $E_1^a \times E_1^b$, $E_2^a \times E_2^b$, ... *each in* $\mathcal{B}\left(S^a \times T^a \times S^b \times T^b\right)$. *Then* $\bigcap_m \mathrm{B}\left(E_m^a \times E_m^b\right) = \mathrm{B}(\bigcap_m \left(E_m^a \times E_m^b\right))$.

Proof. It is immediate from monotonicity that $\mathrm{B}(\bigcap_m \left(E_m^a \times E_m^b\right)) \subseteq \mathrm{B}\left(E_m^a \times E_m^b\right)$, for each m. As such, $\mathrm{B}(\bigcap_m \left(E_m^a \times E_m^b\right)) \subseteq \bigcap_m \mathrm{B}\left(E_m^a \times E_m^b\right)$. We now turn to the opposite inclusion, i.e., $\bigcap_m \mathrm{B}\left(E_m^a \times E_m^b\right) \subseteq \mathrm{B}(\bigcap_m \left(E_m^a \times E_m^b\right))$. Fix a type t^a that believes each E_m^b, i.e., for each $h \in H^a$ and each m, $\beta^a(t^a)\left(E_m^b | S^b(h) \times T^b\right) = 1$. Define $F_m^b = \bigcap_{n=1}^m E_n^b$ and note that, for each h and each (finite) m, $\beta^a(t^a)\left(F_m^b | S^b(h) \times T^b\right) = 1$. Then, for each $h \in H^a$, $\beta^a(t^a)\left(\bigcap_m F_m^b | S^b(h) \times T^b\right) = 1$. (This uses continuity of the probability measure $\beta^a(t^a)\left(\cdot | S^b(h) \times T^b\right)$.) Since, for each $h \in H^a$, $\beta^a(t^a)\left(\bigcap_m E_m^b | S^b(h) \times T^b\right) = \beta^a(t^a)\left(\bigcap_m F_m^b | S^b(h) \times T^b\right) = 1$, t^a believes $\bigcap_m E_m^b$. ■

Definition A3 *Say* $E^a \times E^b \in \mathcal{B}\left(S^a \times T^a \times S^b \times T^b\right)$ *is a* ***self-evident event (in*** $\mathcal{T}$***)*** *if* $E^a \times E^b \subseteq \mathrm{B}\left(E^a \times E^b\right)$.

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 $E^a \times E^b$ obtains and 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.

Fix $E^a \times E^b \in \mathcal{B}\left(S^a \times T^a \times S^b \times T^b\right)$, and iterate the belief operator $\mathrm{B}(\cdot)$: $\mathrm{B}^0\left(E^a \times E^b\right) = E^a \times E^b$ and, for each $m \geq 0$, $\mathrm{B}^{m+1}\left(E^a \times E^b\right) = \mathrm{B}\left(\mathrm{B}^m\left(E^a \times E^b\right)\right)$.

Lemma A1 *Fix an event $E^a \times E^b \in \mathcal{B}\left(S^a \times T^a \times S^b \times T^b\right)$. The following are equivalent:*

(i) $E^a \times E^b$ is self-evident (in $\mathcal{T}$);

(ii) $E^a \times E^b = \bigcap_m \mathrm{B}^m\left(E^a \times E^b\right)$;

(iii) $E^a \times E^b = \bigcap_m \mathrm{B}^m\left(F^a \times F^b\right)$, for some event $F^a \times F^b \in \mathcal{B}\left(S^a \times T^a \times S^b \times T^b\right)$.

Proof. We show that (i) implies (ii). First note, for each event $E^a \times E^b$, $\mathrm{B}^0\left(E^a \times E^b\right) \cap \bigcap_{m\geq 1} \mathrm{B}^m\left(E^a \times E^b\right) \subseteq E$. So, it suffices to show that, if $E^a \times E^b$ is a self-evident event, then $E^a \times E^b \subseteq \mathrm{B}^m\left(E^a \times E^b\right)$, 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., $E^a \times E^b \subseteq \mathrm{B}^m\left(E^a \times E^b\right)$. Then, by monotonicity, $\mathrm{B}\left(E^a \times E^b\right) \subseteq \mathrm{B}\left(\mathrm{B}^m\left(E^a \times E^b\right)\right)$. So, again using the fact that $E^a \times E^b$ is a self-evident event, we have that $E^a \times E^b \subseteq \mathrm{B}\left(E^a \times E^b\right) \subseteq \mathrm{B}^{m+1}\left(E^a \times E^b\right)$.

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 $E^a \times E^b$, $F^a \times F^b \in \mathcal{B}\left(S^a \times T^a \times S^b \times T^b\right)$ with $E^a \times E^b = \bigcap_{m\geq 0} \mathrm{B}^m\left(F^a \times F^b\right)$. Note that

$$\begin{array}{rcl} E^a \times E^b & = & \bigcap_{m\geq 0} \mathrm{B}^m\left(F^a \times F^b\right) \\ & = & \left(F^a \times F^b\right) \cap \left(\bigcap_{m\geq 0} \mathrm{B}\left(\mathrm{B}^m\left(F^a \times F^b\right)\right)\right) \\ & = & \left(F^a \times F^b\right) \cap \mathrm{B}\left(\bigcap_{m\geq 0} \mathrm{B}^m\left(F^a \times F^b\right)\right) \\ & = & \left(F^a \times F^b\right) \cap \mathrm{B}(E^a \times E^b), \end{array}$$

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 $E^a \times E^b \subseteq \mathrm{B}\left(E^a \times E^b\right)$ 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 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 Ω, Φ. Given a measurable map $f : \Omega \to \Phi$, write $\overline{f} : \mathcal{P}(\Omega) \to \mathcal{P}(\Phi)$, for the map where $\overline{f}(\mu)$ is the image measure of μ under f. Note, $\overline{f}$ is measurable. (See Kechris [15, 1995; Exercise 17.40].)

Now, consider two Γ-based type structures, namely $\mathcal{T} = \left\langle S^a, S^b; \mathcal{S}^a, \mathcal{S}^b; T^a, T^b; \beta^a, \beta^b \right\rangle$ and $\mathcal{T}_* = \left\langle S^a, S^b; \mathcal{S}^a_*, \mathcal{S}^b_*; T^a_*, T^b_*; \beta^a_*, \beta^b_* \right\rangle$. 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* $\mu^a \in \mathcal{C}\left(S^b \times T^b; \mathcal{S}^b\right)$. *Define* ν^a *so that, for each* $h \in H^a$, $\overline{(\mathrm{id}^b \times \tau^b)}\left(\mu^a\left(\cdot | S^b(h) \times T^b\right)\right) = \nu^a\left(\cdot | S^b(h) \times T^b_*\right)$. *Then* $\nu^a \in \mathcal{C}\left(S^b \times T^b_*; \mathcal{S}^b_*\right)$.

Proof. It is immediate that ν^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, $(\mathrm{id}^b \times \tau^b)^{-1}(E_*) \in \mathcal{B}\left(S^b \times T^b\right)$. It follows that

$$\begin{aligned}
\nu^a\left(E_* | S^b(i) \times T^b_*\right) &= \mu^a((\mathrm{id}^b \times \tau^b)^{-1}(E_*) | S^b(i) \times T^b) \\
&= \mu^a((\mathrm{id}^b \times \tau^b)^{-1}(E_*) | S^b(h) \times T^b) \times \mu^a\left(S^b(h) \times T^b | S^b(i) \times T^b\right) \\
&= \nu^a\left(E_* | S^b(h) \times T^b_*\right) \times \nu^a\left(S^b(h) \times T^b_* | S^b(i) \times T^b_*\right),
\end{aligned}$$

as required. ■

Given CPS's μ^a and ν^a as in Lemma A2, say **$\boldsymbol{\nu}^{\mathbf{a}}$ is the image CPS of $\boldsymbol{\mu}^{\mathbf{a}}$ under** $\mathrm{id}^{\mathbf{b}} \times \boldsymbol{\tau}^{\mathbf{b}}$. We write $\underline{(\mathrm{id}^b \times \tau^b)} : \mathcal{C}\left(S^b \times T^b; \mathcal{S}^b\right) \to \mathcal{C}\left(S^b \times T^b_*; \mathcal{S}^b_*\right)$ for the associated map, i.e., so that $\underline{(\mathrm{id}^b \times \tau^b)}(\mu^a)$ is the image CPS of μ^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* $\left(\tau^a, \tau^b\right)$ *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* $\underline{(\mathrm{id}^a \times \tau^a)} \circ \beta^b = \beta^b_* \circ \tau^b$.

Given separable metrizable spaces Ω, Φ, 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* $\left(\tau^a, \tau^b\right)$ *a* ***bimeasurable type morphism*** *if it is a type morphism and* τ^a *and* τ^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 Γ*-based type structures* $\mathcal{T} = \left\langle S^a, S^b; \mathcal{S}^a, \mathcal{S}^b; T^a, T^b; \beta^a, \beta^b \right\rangle$ *and* $\mathcal{T}_* = \left\langle S^a, S^b; \mathcal{S}^a_*, \mathcal{S}^b_*; T^a_*, T^b_*; \beta^a_*, \beta^b_* \right\rangle$*. If* $\left(\tau^a, \tau^b\right)$ *is a bimeasurable type morphism from* $\mathcal{T}$ *to* $\mathcal{T}_*$*, then* $S^a \times \tau^a\left(T^a\right) \times S^b \times \tau^b\left(T^b\right)$ *is self-evident in* $\mathcal{T}_*$*.*

(ii) Fix a Γ*-based type structure* $\mathcal{T}_* = \left\langle S^a, S^b; \mathcal{S}^a_*, \mathcal{S}^b_*; T^a_*, T^b_*; \beta^a_*, \beta^b_* \right\rangle$ *and a self-evident event* $S^a \times E^a_* \times S^b \times E^b_* \in \mathcal{B}\left(S^a \times T^a_* \times S^b \times T^b_*\right)$*. Then, there is a type structure* $\mathcal{T} = \left\langle S^a, S^b; \mathcal{S}^a, \mathcal{S}^b; E^a_*, E^b_*; \beta^a, \beta^b \right\rangle$ *and a bimeasurable type morphism from* $\mathcal{T}$ *to* $\mathcal{T}_*$*.*

Proof. Begin with part (i). Fix a bimeasurable type morphism, viz. $\left(\tau^a, \tau^b\right)$. Since the maps τ^a and τ^b are bimeasurable, $S^a \times \tau^a\left(T^a\right) \times S^b \times \tau^a\left(T^a\right)$ is contained in $\mathcal{B}\left(S^a \times T^a_* \times S^b \times T^b_*\right)$. We proceed to show that $S^a \times \tau^a\left(T^a\right) \times S^b \times \tau^b\left(T^b\right) \subseteq \mathrm{B}\left(S^a \times \tau^a\left(T^a\right) \times S^b \times \tau^b\left(T^b\right)\right)$. To show this, it suffices to show that, for each $\tau^a\left(t^a\right) \in \tau^a\left(T^a\right)$ and each information set $h \in H^a$,

$$\beta^a_*\left(\tau^a\left(t^a\right)\right)\left(S^b \times \tau^b\left(T^b\right) | S^b\left(h\right) \times T^b_*\right) = 1.$$

(Again, bimeasurability guarantees that $S^b \times \tau^b\left(T^b\right)$ is Borel in $S^b \times T^b_*$.) But this follows from the definition of a type morphism, since

$$\beta^a_*\left(\tau^a\left(t^a\right)\right)\left(S^b \times \tau^b\left(T^b\right) | S^b\left(h\right) \times T^b_*\right) = \beta^a\left(t^a\right)\left(S^b \times T^b | S^b\left(h\right) \times T^b\right) = 1.$$

Turn to part (ii). Take $T^a = E^a_*$, $T^b = E^b_*$, and endow these sets with the relative topology. (Recall that $S^b \times E^b_* \in \mathcal{B}\left(S^b \times T^b_*\right)$, so that $\mathcal{B}\left(S^b \times E^b_*\right) \subseteq \mathcal{B}\left(S^b \times T^b_*\right)$.) For each $t^a \in E^a_*$, define $\beta^a\left(t^a\right)$ so that, for each $F^b \in \mathcal{B}\left(S^b \times T^b\right)$ and each $h \in H^a$,

$$\beta^a\left(t^a\right)\left(F^b | S^b\left(h\right) \times T^b\right) = \beta^a_*\left(t^a\right)\left(F^b | S^b\left(h\right) \times T^b_*\right).$$

Note, $\beta^a\left(t^a\right)$ 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 $h \in H^a$,

$$\beta^a\left(t^a\right)\left(S^b\left(h\right) \times T^b | S^b\left(h\right) \times T^b\right) = \beta^a_*\left(t^a\right)\left(S^b\left(h\right) \times E^b_* | S^b\left(h\right) \times T^b_*\right) = 1,$$

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 $\mathcal{T} = \left\langle S^a, S^b; \mathcal{S}^a, \mathcal{S}^b; T^a, T^b; \beta^a, \beta^b \right\rangle$ is indeed a Γ-based type structure: It is immediate that T^a and T^b are separable metrizable. So, it suffices to show that β^a and β^b are measurable. We show this for β^a, and an analogous argument establishes the result for β^b.

To show that β^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

$$f : \mathcal{P}(S^b(h)\times T^b) \rightarrow \{\mu \in \mathcal{P}(S^b(h)\times T^b_*) : \mu(S^b(h)\times E^b_*) = 1\}.$$

(See Kechris [15, 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 β^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, take $\tau^a : T^a \rightarrow T^a_*$ and $\tau^b : T^b \rightarrow T^b_*$ to be the identity maps. Certainly, they are bimeasurable. We will show that (τ^a, τ^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. (τ^a, τ^b), from $\mathcal{T}$ to $\mathcal{T}_*$. Let $E_*(\mathcal{T})$ be the self-evident event in $\mathcal{T}_*$ corresponding to $\mathcal{T}$.[12] 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 Γ-based structures $\mathcal{T}$ and $\mathcal{T}_*$, so that there is a bimeasurable type morphism, viz. (τ^a, τ^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 $(s^a, \tau^a(t^a), s^b, \tau^b(t^b)) \in \mathrm{CSB}^m_*((R^a_* \times R^b_*) \cap (S^a \times$

---

[12] That is, $E_*(\mathcal{T}) = S^a \times \tau^a(T^a) \times S^b \times \tau^b(T^b)$.

$\tau^a(T^a) \times S^b \times \tau^b(T^b)))$.

(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. (τ^a, τ^b), from $\mathcal{T}$ to $\mathcal{T}_*$. For the structure $\mathcal{T}$, write $R^{a,0} = R$ 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* Γ*-based structures* $\mathcal{T}$ *and* $\mathcal{T}_*$*, and a type morphism, viz.* (τ^a, τ^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

$$\mathrm{marg}_{S^b(h)}\, \beta^a(t^a)\left(\cdot | S^b(h) \times T^b\right) = \mathrm{marg}_{S^b(h)}\, \beta^a_*(t^a_*)\left(\cdot | S^b(h) \times T^b_*\right),$$

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,

$$\beta^a_*(\tau^a(t^a))\left(E^b \times T^b_* | S^b(h) \times T^b_*\right) = \beta^a(t^a)\left(E^b \times (\tau^b)^{-1}(T^b_*) | S^b(h) \times T^b\right) = \beta^a(t^a)\left(E^b \times T^b | S^b(h) \times T^b\right),$$

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 so $t^b_* \in \tau^b(T^b)$. 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 $R_*^{b,m} \in \mathcal{B}\left(S^b \times T_*^b\right)$. To see this, use Fact I, i.e., $R_*^{b,m} = (\mathrm{id}^b \times \tau^b)\left(R^{b,m}\right)$. Since t^a strongly believes $R^{b,m}$, it follows that $R^{b,m} \in \mathcal{B}\left(S^b \times T^b\right)$. 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 $R_*^{b,m} \cap \left[S^b(h) \times T_*^b\right] \neq \emptyset$. Note that

$$\begin{aligned}
\beta_*^a(\tau^a(t^a))(R_*^{b,m}|S^b(h) \times T_*^b) &= \beta^a(t^a)((\mathrm{id}^b \times \tau^b)^{-1}(R_*^{b,m})|S^b(h) \times T^b) \\
&= \beta^a(t^a)(\{(s^b, t^b) : (s^b, \tau^b(t^b)) \in R_*^{b,m}\}|S^b(h) \times T^b) \\
&= \beta^a(t^a)\left(R^{b,m}|S^b(h) \times T^b\right)
\end{aligned}$$

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, $R^{b,m} \cap \left[S^b(h) \times T^b\right] \neq \emptyset$. So, with the above and the fact that t^a strongly believes $R^{b,m}$,

$$\beta_*^a(\tau^a(t^a))(R_*^{b,m}|S^b(h) \times T_*^b) = \beta^a(t^a)\left(R^{b,m}|S^b(h) \times T^b\right) = 1,$$

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 $R_*^{b,m} \in \mathcal{B}\left(S^b \times T_*^b\right)$. By Facts I-II, plus the observation that $\mathrm{id}^b \times \tau^b$ is measurable, $R^{b,m} = \left(\mathrm{id} \times \tau^b\right)^{-1}(R_*^{b,m})$ is Borel.

Now, fix an information set $h \in H^a$ with $R^{b,m} \cap \left[S^b(h) \times T^b\right] \neq \emptyset$. Note that

$$\begin{aligned}
\beta^a(t^a)\left(R^{b,m}|S^b(h) \times T^b\right) &= \beta^a(t^a)((\mathrm{id}^b \times \tau^b)^{-1}((\mathrm{id}^b \times \tau^b)\left(R^{b,m}\right))|S^b(h) \times T^b) \\
&= \beta_*^a(\tau^a(t^a))((\mathrm{id}^b \times \tau^b)\left(R^{b,m}\right)|S^b(h) \times T_*^b) \\
&= \beta_*^a(\tau^a(t^a))(R_*^{b,m}|S^b(h) \times T_*^b),
\end{aligned}$$

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, $R_*^{b,m} \cap \left[S^b(h) \times T_*^b\right] \neq \emptyset$. So, with the above and the fact that $\tau^a(t^a)$ strongly believes $R_*^{b,m}$,

$$\beta^a(t^a)\left(R^{b,m}|S^b(h) \times T^b\right) = \beta_*^a(\tau^a(t^a))(R_*^{b,m}|S^b(h) \times T_*^b) = 1,$$

as required. ■

III. Self-Evident Events vs. Type Structures. Let us review the approach taken here. We begin with a game Γ, and we will consider the canonical Γ-based type structure, as constructed in

Battigalli-Siniscalchi [5, 1999]. Write $\mathcal{T}_* = \left\langle S^a, S^b; \mathcal{S}^a_*, \mathcal{S}^b_*; T^a_*, T^b_*; \beta^a_*, \beta^b_* \right\rangle$ 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, β^a_* and β^b_* are onto. (See Footnote 2.) Second, $\mathcal{T}_*$ is **terminal**—that is, for each Γ-based structure $\mathcal{T}$, there is a type morphism from $\mathcal{T}$ to $\mathcal{T}_*$.[13]

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 $E^b_* \in \mathcal{B}\left(T^b_*\right)$, so that $S^a \times E^a_* \times S^b \times E^b_*$ is a self evident event, with $S^a \times E^a_* \times S^b \times E^b_* = \bigcap_m \mathrm{B}^m \left(F^a_* \times F^b_*\right)$. Certainly, each $t^b_* \in E^b_*$ believes $\{Up\} \times T^a_*$. Moreover, for each CPS $\mu^a \in \mathcal{C}\left(S^b \times T^b_*\right)$ (resp. $\mu^b \in \mathcal{C}\left(S^a \times T^a_*\right)$) 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. $\beta^b_*\left(t^b_*\right) = \mu^b$). (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 Γ-based structure $\mathcal{T}$ to 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 review.

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 [21, 1966], Proposition 3.3.7 in Srivastava [25], and the fact that the map from types to hierarchies is measurable.[14]) 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

---

[13] The terminology is due to Böge-Eisele [10, 1979]. Battigalli-Siniscalchi [5, 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.

[14] See the Online Appendix, on this latter fact.

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, in the canonical construction, no two types induce the same hierarchies of beliefs.

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 $\mu(\cdot|\cdot)$ 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 $(F \times \Omega_2) \cap E \neq \emptyset$. Note that $\operatorname{marg}_{\Omega_1} \mu\left(\operatorname{proj}_{\Omega_1} E|F\right)$ is well defined because $\operatorname{proj}_{\Omega_1} E$ is Borel by assumption. Since μ strongly believes E, $\mu(E|F \times \Omega_2) = 1$. Then certainly $\operatorname{marg}_{\Omega_1} \mu\left(\operatorname{proj}_{\Omega_1} E|F\right) = 1$, as required. ■

Appendix C Proofs for Section 5

In what follows, we fix a set of first-order beliefs $F = F^a \times F^b$, with $F^a \subseteq \mathcal{C}\left(S^b\right)$, $F^b \subseteq \mathcal{C}\left(S^a\right)$.

Lemma C1 *Fix $s^a \in S_F^{a,m+1}$, for $m \geq 0$. There exists a CPS $\mu^a(\cdot|\cdot)$ so that $s^a \in \rho^a(\mu^a(\cdot|\cdot))$ and $\mu^a(\cdot|\cdot)$ strongly believes each $S_F^{b,n}$ for $n \leq m$.*

Proof. Fix $s^a \in S_F^{a,m+1}$. Then, for each $n \leq m$, there exists a CPS μ_n^a so that $s^a \in \rho^a(\mu_n^a(\cdot|\cdot))$ and $\mu_n^a(\cdot|\cdot)$ strongly believes $S_F^{b,n}$. We will show that there exists a CPS μ^a so that $s^a \in \rho^a(\mu^a(\cdot|\cdot))$ and $\mu^a(\cdot|\cdot)$ strongly believes each $S_F^{b,n}$ for each $n \leq m$.

Begin by constructing a CPS $\mu^a(\cdot|\cdot)$: For each information set h, set $\mu^a\left(\cdot|S^b(h)\right) = \mu^{a,n}\left(\cdot|S^b(h)\right)$ where $n = \max\{o : S^b(h) \cap S_F^{b,o} \neq \emptyset\}$. To see that $\mu^a(\cdot|\cdot)$ is indeed a CPS, first note that conditions (i)-(ii) are immediate from the construction. For condition (iii), fix an events $E \subseteq S^b(h) \subseteq S^b(i)$. Note, there is some $n \leq m$ so that $\mu^a\left(E|S^b(i)\right) = \mu^{a,n}\left(E|S^b(i)\right)$. For this n, $\mu^{a,n}\left(E|S^b(i)\right) = \mu^{a,n}\left(E|S^b(h)\right)\mu^{a,n}\left(S^b(h)|S^b(i)\right)$, by condition (iii) of a CPS. If $\mu^a\left(\cdot|S^b(h)\right) = \mu^{a,n}\left(\cdot|S^b(h)\right)$, then condition (iii) is immediate. If $\mu^a\left(\cdot|S^b(h)\right) \neq \mu^{a,n}\left(\cdot|S^b(h)\right)$, then $S^b(h) \cap S_F^{b,n} = \emptyset$. By strong belief, $\mu^{a,n}(S_F^{b,n}|S^b(i)) = 1$. So, $\mu^{a,n}\left(S^b(h)|S^b(i)\right) = 0$. Using the fact that $E \subseteq S^b(h) \subseteq S^b(i)$, we then have

$$\begin{aligned}
\mu^a\left(E|S^b(i)\right) &= \mu^{a,n}\left(E|S^b(i)\right) \\
&= 0 \\
&= \mu^a\left(E|S^b(h)\right)\mu^{a,n}\left(S^b(h)|S^b(i)\right) \\
&= \mu^a\left(E|S^b(h)\right)\mu^a\left(S^b(h)|S^b(i)\right),
\end{aligned}$$

as required.

It is immediate from the construction that s^a is sequentially optimal under $\mu^a(\cdot|\cdot)$. (Simply use the fact that s^a is sequentially optimal under each $\mu^{a,n}(\cdot|\cdot)$.) Next, if $S^b(h) \cap S_F^{b,n} \neq \emptyset$, then there exists some $o \geq n$ with $S^b(h) \cap S_F^{b,o} \neq \emptyset$ and $\mu^{a,o}\left(\cdot|S^b(h)\right) = \mu^a\left(\cdot|S^b(h)\right)$. Using the fact that $\mu^{a,o}$ strongly believes $S_F^{b,o}$, $\mu^a(S_F^{b,o}|S^b(h)) = 1$. Since $S_F^{b,o} \subseteq S_F^{b,n}$, we also have that $\mu^a(S_F^{b,n}|S^b(h)) = 1$, as required. ■

Appendix D Proofs for Section 6

In this appendix, we prove 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 $\mu^a(\cdot|\cdot)$ on S^b, so that $s^a \in \rho^a(\mu^a(\cdot|\cdot))$ and $\mu^a(\cdot|\cdot)$ 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 Γ and a subgame Δ. Write H_Δ^a for the set of information sets that are contained in Δ. We will abuse notation and write $S^a(\Delta)$ for the set of strategies of Γ that allow Δ. We also write S_Δ^a for the set of strategies on subgame Δ. Note, each strategy $s_\Delta^a \in S_\Delta^a$ can be viewed as the projection of a strategy $s^a \in S^a(\Delta)$ into $\prod_{h \in H_\Delta^a} C^a(h)$. Given a set $E^a \subseteq S^a$, write E_Δ^a for the set of strategies $s_\Delta^a \in S_\Delta^a$ so that there is some $s^a \in E^a \cap S^a(\Delta)$ whose projection into $\prod_{h \in H_\Delta^a} C^a(h)$ is s_Δ^a. We will write π_Δ^a and π_Δ^b for the payoff functions associated with the subtree Δ. So, if $\left(s^a, s^b\right)$ allows Δ, then $\pi_\Delta\left(s_\Delta^a, s_\Delta^b\right) = \pi\left(s^a, s^b\right)$.

Lemma D1 *Fix a game Γ and a subgame Δ. If $Q^a \times Q^b$ satisfies the best response property for the game Γ, then $Q_\Delta^a \times Q_\Delta^b$ satisfies the best response property for the subgame Δ.*

Proof. If $Q_\Delta^a \times Q_\Delta^b = \emptyset$, then it is immediate that $Q_\Delta^a \times Q_\Delta^b$ satisfies the best response property. So, we will suppose $Q_\Delta^a \times Q_\Delta^b \neq \emptyset$.

Fix a strategy $s_\Delta^a \in Q_\Delta^a$. Then there exists a strategy $s^a \in Q^a \cap S^a(\Delta)$ whose projection into $\prod_{h \in H_\Delta^a} C^a(h)$ is s_Δ^a. Since $s^a \in Q^a$, we can find a CPS $\mu^a(\cdot|\cdot) : \mathcal{B}\left(S^b\right) \times \mathcal{S}^b \to [0,1]$ so that $s^a \in \rho^a(\mu^a(\cdot|\cdot))$ and $\mu^a(\cdot|\cdot)$ strongly believes Q^b.

Let $\mathcal{S}_\Delta^b$ be the set of all $S_\Delta^b(h)$ for $h \in H_\Delta^a$. Define $\nu_\Delta^a(\cdot|\cdot) : \mathcal{B}\left(S_\Delta^b\right) \times \mathcal{S}_\Delta^b \to [0,1]$ so that, for each event $E^b \subset S^b$ and each $S_\Delta^b(h) \in \mathcal{S}_\Delta^b$, $\nu_\Delta^a\left(E_\Delta^b|S_\Delta^b(h)\right) = \mu^a\left(E^b|S^b(h)\right)$. It is readily verified that $\nu_\Delta^a(\cdot|\cdot)$ is indeed a CPS on $\left(S_\Delta^b, \mathcal{S}_\Delta^b\right)$.

Since s^a allows Δ and s^a is sequentially optimal under μ^a, it follows that s^a_Δ is sequentially optimal under ν^a_Δ. Fix some $S^b_\Delta(h) \in \mathcal{S}^b_\Delta$. If $Q^b_\Delta \cap S^b_\Delta(h) \neq \emptyset$, then $Q^b \cap S^b(h) \neq \emptyset$. So, in this case, $\nu^a_\Delta\left(Q^b_\Delta | S^b_\Delta(h)\right) = \mu^a\left(Q^b | S^b(h)\right) = 1$. This establishes that ν^a_Δ strongly believes Q^b_Δ.

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 [12, 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, $\pi^a\left(s^a, s^b\right) = c$ if s^b plays Cooperate in the first period, and $\pi^a\left(s^a, s^b\right) = e$ 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 $\pi^a\left(r^a, s^b\right) \geq d$ and, if $s^b \in Q^b$ plays Defect in the first period, then $\pi^a\left(r^a, s^b\right) \geq 0$.

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.*

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 $\left(s^a, s^b\right) \in Q^a \times Q^b$ is outcome equivalent to a Nash Equilibrium.*

The proof of this Lemma closely follows the proof of Proposition 6.1a in Brandenburger-Friedenberg [12, 2004]. It is by induction on the length of the tree. Specifically, fix a game Γ and a subgame Δ. The induction hypothesis states that if a set satisfies the best response property on Δ 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 Γ, it also satisfies the best response property on the subgame Δ. (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* $\left(s^a, s^b\right)$, $\left(r^a, r^b\right) \in Q^a \times Q^b$, $\pi\left(s^a, s^b\right) = \pi\left(r^a, r^b\right)$.

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 $\left(s^a, s^b\right)$ and $\left(r^a, s^b\right)$ 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 $\Delta = k$.

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 $s^a, r^a \in Q^a \cap S^a(\Delta)$ and $s^b, r^b \in Q^b$, we have that $\pi\left(s^a, s^b\right) = \pi\left(r^a, r^b\right)$.

Now, fix some $\left(s^a, s^b\right), \left(r^a, r^b\right) \in Q^a \times Q^b$, where $s^a \in S^a(k)$ and $r^a \in S^a(j)$. We have already established that $\pi\left(s^a, s^b\right) = \pi\left(r^a, r^b\right)$, 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 $\mu^a\left(\cdot|S^b\right)$ with $\mu^a\left(Q^b|S^b\right) = 1$. With this,

$$\begin{aligned}
\pi^a\left(s^a, s^b\right) &= \textstyle\sum_{q^b \in Q^b} \pi^a\left(s^a, q^b\right) \mu^a\left(q^b|S^b\right) \\
&\geq \textstyle\sum_{q^b \in Q^b} \pi^a\left(r^a, q^b\right) \mu^a\left(q^b|S^b\right) \\
&= \pi^a\left(r^a, r^b\right).
\end{aligned}$$

(The first equality follows from the fact that, for each $q^b \in Q^b$, $\pi^a\left(s^a, s^b\right) = \pi^a\left(s^a, q^b\right)$. This is a consequence of the last line in the preceding paragraph. Likewise, for the last equality.) By an analogous argument, $\pi^a\left(r^a, r^b\right) \geq \pi^a\left(s^a, s^b\right)$. So, $\pi^a\left(r^a, r^b\right) = \pi^a\left(s^a, s^b\right)$. Using the single payoff condition, $\pi^b\left(r^a, r^b\right) = \pi^b\left(s^a, s^b\right)$. ■

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 $\Delta = k$.

Fix some $\left(s^a, s^b\right) \in Q^a \times Q^b$ and suppose $s^a \in S^a(1)$. Note, $Q_1^a \times Q_1^b$ satisfies the best response property (Lemma D1). So, by the induction hypothesis, there is a Nash equilibrium of subgame 1, viz. $\left(r_1^a, r_1^b\right)$, so that $\pi\left(s_1^a, s_1^b\right) = \pi\left(r_1^a, r_1^b\right)$. Consider a strategy $r^a \in S^a(1)$ so that the projection of r^a onto $\prod_{h \in H_1^a} C^a(h)$ is r_1^a. We need to show that we can choose $r_2^b, \ldots, r_K^b \in \times_{k=2}^K S_k^b$ so that, for each $q^a \in Q^a$ and associated $q_k^a \in S_k^a$, $\pi^a\left(r_1^a, r_1^b\right) \geq \pi^a\left(q_k^a, r_k^b\right)$. The profile $\left(r^a, \left(r_1^b, r_2^b, \ldots, r_K^b\right)\right)$ will then be a Nash Equilibrium of the game.

Since $s^a \in Q^a$, there exists a CPS and an associated measure $\mu^a\left(\cdot|S^b\right)$ so that

$$\sum\nolimits_{s^b \in S^b}\left[\pi^a\left(s^a, s^b\right) - \pi^a\left(q^a, s^b\right)\right]\mu^a\left(s^b|S^b\right) \geq 0,$$

for all $q^a \in S^a$. Fix k from $2, \ldots, K$. Using Lemma E2,

$$\pi^a\left(r_1^a, r_1^b\right) = \pi^a\left(s_1^a, s_1^b\right) \geq \sum\nolimits_{s_k^b \in S_k^b}\pi^a\left(q_k^a, s_k^b\right)\left(\operatorname{marg}_{S_k^b}\mu\left(\cdot|S^b\right)\right)\left(s_k^b\right),$$

for any $q_k^a \in S_k^a$. Letting $\left(\overline{q}_k^a, \overline{q}_k^b\right) \in \arg\max_{S_k^a}\min_{S_k^b}\pi^a(\cdot,\cdot)$, we have in particular

$$\pi^a\left(r_1^a, r_1^b\right) \geq \sum\nolimits_{s_k^b \in S_k^b}\pi^a\left(\overline{q}_k^a, s_k^b\right)\left(\operatorname{marg}_{S_k^b}\mu\left(\cdot|S^b\right)\right)\left(s_k^b\right).$$

But $\pi^a\left(\overline{q}_k^a, q_k^b\right) \geq \pi^a\left(\overline{q}_k^a, \overline{q}_k^b\right)$ for any $q_k^b \in S_k^b$, by definition. So

$$\pi^a\left(r_1^a, r_1^b\right) \geq \sum\nolimits_{s_k^b \in S_k^b}\pi^a\left(\overline{q}_k^a, \overline{q}_k^b\right)\left(\operatorname{marg}_{S_k^b}\mu\left(\cdot|S^b\right)\right)\left(s_k^b\right) = \pi^a\left(\overline{q}_k^a, \overline{q}_k^b\right).$$

Set $(\underline{q}_k^a, \underline{q}_k^b) \in \arg\min_{S_k^b}\max_{S_k^a}\pi^a(\cdot,\cdot)$. By the Minimax Theorem for PI games (see, e.g., Ben Porath [9, 1997]), $\pi^a\left(\overline{q}_k^a, \overline{q}_k^b\right) = \pi^a(\underline{q}_k^a, \underline{q}_k^b)$. It follows that $\pi^a(r_1^a, r_1^b) \geq \pi^a\left(\overline{q}_k^a, \overline{q}_k^b\right) = \pi^a(\underline{q}_k^a, \underline{q}_k^b)$. But $\pi^a(\underline{q}_k^a, \underline{q}_k^b) \geq \pi^a(q_k^a, \underline{q}_k^b)$ for any $q_k^a \in S_k^a$, by definition. So $\pi^a(r_1^a, r_1^b) \geq \pi^a(q_k^a, \underline{q}_k^b)$, for each $q_k^a \in S_k^a$. Setting each $r_k^b = \underline{q}_k^b$ 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 $Q^a \times Q^b = \left\{\left(s^a, s^b\right)\right\}$, where $\left(s^a, s^b\right)$ 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 $\mu^a(\cdot|\cdot)$, 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 $\mu^a(\cdot|\cdot)$. We use the fact that $\mu^a(\cdot|\cdot)$ 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 $\mu^a(\cdot|\cdot)$ (resp. $\mu^b(\cdot|\cdot)$) 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 $\mu^a(\cdot|\cdot)$ (resp. $\mu^b(\cdot|\cdot)$) 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* Ω *is finite. Let* $\mu(\cdot|\cdot)$ *be a CPS on* $(\Omega, \mathcal{E})$ *and let* ϖ *be a measure on* Ω. *Construct* $\nu(\cdot|\cdot) : \mathcal{B}(\Omega) \times \mathcal{E} \to [0,1]$ *as follows: If* $F \in \mathcal{E}$ *with* $\operatorname{Supp} \varpi \cap F \neq \emptyset$ *then* $\nu(\cdot|F) = \varpi(\cdot|F)$. *Otherwise,* $\nu(\cdot|F) = \mu(\cdot|F)$. *Then* $\nu(\cdot|\cdot)$ *is a CPS.*

Proof. Let $\mu(\cdot|\cdot)$, ϖ, and $\nu(\cdot|\cdot)$ 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 $\operatorname{Supp} \varpi \cap F \neq \emptyset$. 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 $\operatorname{Supp} \varpi \cap G = \emptyset$. Then $\operatorname{Supp} \varpi \cap F = \emptyset$, 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 $\operatorname{Supp} \varpi \cap F = \emptyset$ but $\operatorname{Supp} \varpi \cap G \neq \emptyset$. Then

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

where the last equality follows from the fact that $\operatorname{Supp} \varpi \cap F = \emptyset$. 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 $\mu^a(\cdot|\cdot)$ on $(S^b, \mathcal{S}^b)$ so that:*

(i) $Q^a \subseteq P^a = \rho^a(\mu^a(\cdot|\cdot))$,

(ii) $\mu^a(\cdot|\cdot)$ 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 $\mu^a(\cdot|\cdot)$ and $\mu^b(\cdot|\cdot)$ 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 $\mu^a(\cdot|\cdot)$ and $\mu^b(\cdot|\cdot)$. 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* $\mu^a \in \mathcal{P}\left(S^b\right)$ *so that* s^a *is optimal under* μ^a *given* S^a*. Then, for any information set* h *with* $s^a \in S^a(h)$ *and* $\mu^a\left(S^b(h)\right) > 0$, s^a *is optimal under* $\mu^a\left(\cdot|S^b(h)\right)$ *given* $S^a(h)$.

Given a measure $\mu \in \mathcal{P}(\Omega)$, we write $\operatorname{Supp}\mu$ 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 $\mu^a(\cdot|\cdot)$ (resp. $\mu^b(\cdot|\cdot)$) satisfying the conditions of Definition E2. Let X^a (resp. X^b) be the set of strategies that are optimal under $\mu^a\left(\cdot|S^b\right)$ (resp. $\mu^b(\cdot|S^a)$). By Lemma E5, $X^a \times X^b$ is a constant set.

Construct a measure $\varpi^a \in \mathcal{P}\left(S^b\right)$ as follows: Begin with a measure $\overline{\varpi}^a$ with $\operatorname{Supp}\overline{\varpi}^a = P^b$. Construct ϖ^b so that, for each $r^b \in P^b$,

$$\varpi^a\left(r^b\right) = (1-\varepsilon)\,\mu^a\left(r^b|S^b\right) + \varepsilon\overline{\varpi}^a\left(r^b\right),$$

where $\varepsilon \in (0,1)$. Note that μ^a strongly believes $Q^b \subseteq P^b$, $\operatorname{Supp}\mu^a\left(\cdot|S^b\right) \subseteq P^b$. 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 ε 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 ϖ^a so that s^a is optimal under ϖ^a if and only if $s^a \in X^a$.

Now construct a CPS $\nu^a(\cdot|\cdot)$ on $\left(S^b, \mathcal{S}^b\right)$ as follows: If $P^b \cap S^b(h) \neq \emptyset$, let $\nu^a\left(\cdot|S^b(h)\right) = \varpi^a\left(\cdot|S^b(h)\right)$. (This is well defined since, in this case, $\varpi^a\left(S^b(h)\right) > 0$.) If $P^b \cap S^b(h) = \emptyset$, let $\nu^a\left(\cdot|S^b(h)\right) = \mu^a\left(\cdot|S^b(h)\right)$. Lemma E3 establishes that $\nu^a(\cdot|\cdot)$ is a CPS. Construct a measure ϖ^b on S^a and a CPS $\nu^b(\cdot|\cdot)$ on $(S^a, \mathcal{S}^a)$ analogously.

Take $W^a = \rho^a(\nu^a(\cdot|\cdot))$ and $W^b = \rho^b(\nu^b(\cdot|\cdot))$. 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, $W^a = \rho^a(\nu^a(\cdot|\cdot))$. 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 ϖ^a. Let $h \in H^a$ with $s^a \in S^a(h)$. If $P^b \cap S^b(h) \neq \emptyset$ then $\varpi^a\left(\cdot|S^b(h)\right) = \nu^a\left(\cdot|S^b(h)\right)$ and s^a is optimal under $\nu^a\left(\cdot|S^b(h)\right)$ among all strategies in $S^a(h)$. (See Lemma E6.) If $P^b \cap S^b(h) = \emptyset$ then $\nu^a\left(\cdot|S^b(h)\right) = \mu^a\left(\cdot|S^b(h)\right)$. So, again, s^a is optimal under $\nu^a\left(\cdot|S^b(h)\right)$ 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 $\nu^a(\cdot|\cdot)$ strongly believes P^b. For this notice that if $P^b \cap S^b(h) \neq \emptyset$ then $\nu^a\left(P^b|S^b(h)\right) = \varpi^a\left(P^b|S^b(h)\right) = 1$.

Finally, we show condition (iii). Suppose r^a is optimal under $\nu^a\left(\cdot|S^b\right)$. We will show that $\pi^a\left(r^a, s^b\right) = \pi^a\left(s^a, s^b\right)$ for all $\left(s^a, s^b\right) \in P^a \times P^b$. To see this, recall, $\nu^a\left(\cdot|S^b\right) = \varpi^a$. So, if r^a is optimal under $\nu^a\left(\cdot|S^b\right)$ 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 $\left(s^a, s^b\right)$ be a Nash equilibrium in sequentially justifiable strategies. Then there exists an EFBRS, viz. $Q^a \times Q^b$, that contains $\left(s^a, s^b\right)$.*

Proof. Fix a Nash equilibrium in sequentially optimal strategies, viz. $\left(s^a, s^b\right)$. Let $Q_0^a \times Q_0^b = \left\{s^a\right\} \times \left\{s^b\right\}$. By Lemma E4, $Q_0^a \times Q_0^b$ satisfies the best response property. So, there is a CPS $\mu^a(\cdot|\cdot)$ (resp. $\mu^b(\cdot|\cdot)$) that strongly believes $\left\{s^b\right\}$ (resp. $\{s^a\}$) and s^a (resp. s^b) is sequentially optimal under $\mu^a(\cdot|\cdot)$ (resp. $\mu^b(\cdot|\cdot)$). Let $Q_1^a = \rho^a(\mu^a(\cdot|\cdot))$ (resp. $Q_1^b = \rho^b(\mu^a(\cdot|\cdot))$). Note that $Q_1^a \times Q_1^b$ is an expansion of $Q_0^a \times Q_0^b$ (associated with the CPS's $\mu^a(\cdot|\cdot)$ and $\mu^b(\cdot|\cdot)$). Now, repeatedly apply Lemma E7 to get sets $Q_0^a \times Q_0^b$, $Q_1^a \times Q_1^b$, $Q_2^a \times Q_2^b$, …, where each $Q_{m+1}^a \times Q_{m+1}^b$ is an expansion of $Q_m^a \times Q_m^b$. Since the game is finite, there is some M with $Q_m^a \times Q_m^b = Q_M^a \times Q_M^b$ for all $m \geq M$. The set $Q_M^a \times Q_M^b$ 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 $\left(s^a, s^b\right) \in Q^a \times Q^b$ 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. $\left(r_1^a, r_1^b\right)$. Following the proof, we use this, to construct a Nash equilibrium $(r^a, (r_1^b, \underline{q}_2^b, \ldots, \underline{q}_K^b))$, where each $\underline{q}_k^b$ 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 $\underline{q}_k^b$ (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.

Let us consider one related result, which may shed light on the gap. Consider a pure strategy profile $\left(s^a, s^b\right)$ and a mixed strategy profile $\left(\sigma^a, \sigma^b\right) \in \mathcal{P}(S^a) \times \mathcal{P}\left(S^b\right)$. Call $\left(s^a, s^b\right)$ and $\left(\sigma^a, \sigma^b\right)$ outcome equivalent if $\pi\left(s^a, s^b\right) = \pi\left(\sigma^a, \sigma^b\right)$. Likewise, call $Q^a \times Q^b \subseteq S^a \times S^b$ and $\left(\sigma^a, \sigma^b\right) \in \mathcal{P}(S^a) \times \mathcal{P}\left(S^b\right)$ outcome equivalent if each $\left(s^a, s^b\right) \in Q^a \times Q^b$ is outcome equivalent to $\left(\sigma^a, \sigma^b\right)$. Then:

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. $\left(\sigma^a, \sigma^b\right)$, so that:*

(i) $Q^a \times Q^b$ *is outcome equivalent to* $\left(\sigma^a, \sigma^b\right)$*, and*

(ii) each $s^a \in \operatorname{Supp} \sigma^a$ *(resp.* $s^b \in \operatorname{Supp} \sigma^b$*) is sequentially justifiable.*

Proof. Pick some $\left(r^a, r^b\right) \in Q^a \times Q^b$ and let $\mu^a(\cdot|\cdot)$ be a CPS so that $r^a \in \rho^a\left(\mu^a(\cdot|\cdot)\right)$ and $\mu^a(\cdot|\cdot)$ strongly believes Q^b. Set $\sigma^b = \mu^a\left(\cdot|S^b\right)$. Construct σ^a analogously.

First, notice that $\left(\sigma^a, \sigma^b\right)$ is a mixed strategy Nash equilibrium: Begin by using the fact that $\mu^b\left(Q^a|S^a\right) = 1$ and $\mu^a\left(Q^b|S^b\right) = 1$. 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 $\left(s^a, s^b\right) \in \operatorname{Supp} \sigma^a \times \operatorname{Supp} \sigma^b$, $\pi\left(s^a, s^b\right) = \pi\left(r^a, r^b\right)$. So, for each $s^a \in \operatorname{Supp} \sigma^a$ and each $q^a \in S^a$,

$$\begin{aligned} \pi^a\left(s^a, \sigma^b\right) &= \pi^a\left(r^a, r^b\right) \\ &= \pi^a\left(r^a, \sigma^b\right) \geq \pi^a\left(q^a, \sigma^b\right), \end{aligned}$$

where the inequality holds because $r^a \in \rho^a\left(\mu^a(\cdot|\cdot)\right)$ and $\mu^a\left(\cdot|S^b\right) = \sigma^b$. Applying an analogous argument to b, establishes that $\left(\sigma^a, \sigma^b\right)$ is indeed a Nash equilibrium.

Next, notice that $Q^a \times Q^b$ is outcome equivalent to $\left(\sigma^a, \sigma^b\right)$: 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 $\left(s^a, s^b\right) \in Q^a \times Q^b$, $\pi\left(s^a, s^b\right) = \pi\left(\sigma^a, \sigma^b\right)$.

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. ■

Putting Lemmata E2-E9 together, we get the following corollary:

Corollary E2 *Fix a PI game satisfying SPC. If* $Q^a \times Q^b$ *satisfies the best response property, then there exists a mixed strategy Nash equilibrium, viz.* $\left(\sigma^a, \sigma^b\right)$*, so that:*

(i) $Q^a \times Q^b$ *is outcome equivalent to* $\left(\sigma^a, \sigma^b\right)$*, and*

(ii) each $s^a \in \operatorname{Supp} \sigma^a$ *(resp.* $s^b \in \operatorname{Supp} \sigma^b$*) is sequentially justifiable.*

Fix a perfect-information SPC game. Corollary E2 gives that, if we begin with a set satisfying the best response property, 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 give that the mixed strategy itself is sequentially justifiable. Indeed, for non-PI games, we can construct a Nash equilibrium per Lemma E9, where the mixed strategy is not sequentially justifiable. It remains open whether the same can be done in a perfect-information SPC game.

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., and A. Prestipino, “Transparent Restrictions on Beliefs and Forward Induction Reasoning in Games with Payoff Uncertainty,” 2009.

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

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

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

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

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

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

[11] 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.

[12] Brandenburger, A., and A. Friedenberg, “Self-Admissible Sets,” unpublished, 2004, available at www.stern.nyu.edu/∼abranden.

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

[14] Hillas, J. and E. Kohlberg, “Foundations of Strategic Equilibrium,” TBA

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

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

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

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

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

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

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

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

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

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

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

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

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