Local reasoning in dynamic games*

EPICENTER Working Paper No. 4 (2015)

ANDRÉS PEREA[†]

Department of Quantitative Economics & EpiCenter, Maastricht University

ELIAS TSAKAS[‡]

Department of Economics & EpiCenter, Maastricht University

January 26, 2015

Abstract

In this paper we introduce a novel framework for modeling the players' reasoning in a dynamic game: at each history each active player reasons about *her opponents' rationality at certain histories only*. As a result we obtain a generalized solution concept, called local common strong belief in rationality, and we characterize the strategy profiles that can be rationally played under our concept by means of a simple elimination procedure. Finally, we show that standard models of reasoning can be embedded as special cases on our framework. In particular, the forward induction concept of common strong belief in rationality (Battigalli and Siniscalchi, 2002) is a special case of our model with the players reasoning about all histories, whereas the backward induction concept of common belief in future rationality (Perea, 2014) is a special case of our model with the players reasoning about future histories only.

KEYWORDS: Local strong belief in rationality, forward induction, backward induction, local iterated conditional dominance procedure.

*We are indebted to the audiences in Maastricht, Bilbao, Brisbane, LogiCIC Workshop (Amsterdam), SAET (Paris) and LOFT (Bergen) for valuable comments and useful discussions.

[†]Department of Quantitative Economics & EpiCenter, Maastricht University, P.O. Box 616, 6200 MD, Maastricht, The Netherlands; Tel: +31-43-38 83922; E-mail: `a.perea@maastrichtuniversity.nl`

[‡]Department of Economics (AE1) & EpiCenter, Maastricht University, P.O. Box 616, 6200 MD, Maastricht, The Netherlands; Tel: +31-43-38 83649; E-mail: `e.tsakas@maastrichtuniversity.nl`

1. Introduction

The two standard families of solution concepts for dynamic games are forward induction (FI) and backward induction (BI). The main difference between the two is that while FI solution concepts explicitly or implicitly assume that players use information from past observations to assess the opponents' rationality in the future, BI concepts on the other hand typically postulate that players believe at every history that their opponents will play rationally from that point onwards irrespective of how they have behaved so far.[1] Well-known examples of FI include extensive-form rationalizability (Pearce, 1984) and extensive-from best response sets (Battigalli and Friedenberg, 2012). On the other hand, BI contains concepts like subgame perfect equilibrium (Selten, 1965), sequential equilibrium (Kreps and Wilson, 1982), backward dominance procedure (Perea, 2014) and backward rationalizability (Penta, 2009).[2]

With the recent surge of the epistemic approach to game theory, most of the aforementioned concepts have been characterized in terms of conditions on the players' beliefs about the opponents' rationality. In particular, epistemic game theory provides a framework to formally incorporate belief hierarchies in our game-theoretic analysis,[3] and consequently it allows us to express conditions such as belief in the opponents' rationality. However, note that specifically in dynamic games, the interpretation of "belief in the opponents' rationality" is not as straightforward as it is in static games. The reason is precisely the dynamic nature of the game, i.e., the fact that players may move more than once throughout the game. To see this consider a game with Ann and Bob moving alternatingly, and assume that Bob observes Ann having chosen a strictly dominated action in the previous period. Then, can we say that Bob believes in Ann's rationality? The answer to this question actually depends on the notion of "belief in the opponents' rationality" that we employ. The two main notions of belief in the opponents' rationality in the literature are strong belief in rationality (Battigalli and Siniscalchi, 2002) and belief in future rationality (Perea, 2014).

We say that Bob strongly believes in Ann's rationality at some history h whenever the following holds: if Bob is able to rationalize Ann's moves at all histories leading to h then he believes that Ann will behave rationally from h onwards. In other words, strong belief in rationality postulates that players look into their opponents' past behavior in order to assess whether their opponents will act rationally in future. This already suggests that strong belief in rationality may be the appropriate

[1]We should make clear that the standard backward induction procedure is merely a solution concept within the family of BI concepts, and it is formally defined only for extensive-form games with perfect information and without relevant ties.

[2]For an overview of this literature we refer to the textbook by Perea (2012).

[3]In a dynamic game, a (conditional) belief hierarchy describes what each player at each history believes about what each opponent at each history will do, and also what each player at each history believes about what each opponent at each history believes about what each opponent at each history will do, and so on (Battigalli and Siniscalchi, 1999).

tool for characterizing FI solution concepts, which as we have already mentioned incorporate the idea that players use information from observed behavior to form beliefs about their opponents' future rationality. Indeed, this is the case, e.g., extensive-form rationalizability is epistemically characterized by rationality and common strong belief in rationality (Battigalli and Siniscalchi, 2002).[4]

We say that Bob believes in Ann's future rationality at some history h whenever he believes that Ann will behave rationally from h onwards irrespective of how she has behaved at the histories leading to h. That is, belief in the opponents' future rationality postulates that players disregard their opponents' past behavior when they assess whether their opponents will act rationally in the future. This suggests that belief in the opponents' future rationality is perhaps the appropriate condition for characterizing BI solution concepts, which as we have previously mentioned incorporate the idea that players do not look into the past. It turns out that this is indeed the case, e.g., both backward rationalizability and the backward dominance procedure are epistemically characterized by rationality and common belief in future rationality (Perea, 2014).

From the previous preliminary analysis, it already becomes clear that the main difference between FI and BI is the extent to which players reason about past histories. This already suggests that the two may be embedded as special cases of a generalized parametric solution concept, which is what we do in this paper. The main idea is that each player i reasons at each history h only about an exogenously specified set of histories, henceforth denoted by $F_i(h)$.[5] We call this type of reasoning local strong belief in rationality. Formally, we say that Bob locally strongly believes in Ann's rationality at h whenever the following holds: if Bob is able to rationalize at h Ann's moves at every history in $F_{\mathrm{Bob}}(h)$ then he believes that Ann is indeed rational at every history in $F_{\mathrm{Bob}}(h)$. The exact specification of F_i determines whether local strong belief in rationality is a FI or a BI concept or perhaps none of these. This unification into in a generalized framework allows us to deeply understand the fundamental similarities/differences between FI and BI reasoning. In fact, there are already several results relating FI and BI in terms of predicted outcomes (Battigalli, 1997; Chen and Micali, 2013; Heifetz and Perea, 2014), but still the intuitive relationship of the two remains quite unclear. In this sense, this paper constitutes the first systematic attempt to close this conceptual gap. Last but not least, this framework would also allow us to study other interesting special cases that have not appeared in the literature so far, such as for instances cases where the players reason only about some particular focal histories.

The first general result in the paper provides a simple iterative procedure that induces the strate-

[4]In the formal statement of the result, Battigalli and Siniscalchi (2002) require a complete type space model. In fact, without the completeness assumption rationality and common strong belief in rationality lead to another FI concept, namely to an extensive-form best response set (Battigalli and Friedenberg, 2012).

[5]In fact, we introduce a function F_i mapping each history h where player i is active to a collection of histories where at least one of the opponents is active.

gies that can be rationally played by each player at each history under our solution concept for an arbitrary specification of F_i. The procedure bears several similarities with the iterated conditional dominance procedure which yields the strategies that can be rationally played under common strong belief in rationality (Shimoji and Watson, 1998), as well as with the backward dominance procedure which yields the strategies that can be rationally played under common belief in future rationality (Perea, 2014). Our procedure simultaneously eliminates strategies and conditional beliefs for each history at each step, thus inducing not only the predictions of our concepts, but also the outcome of the reasoning of each player at each history.

Finally, as we have already mentioned above, we formally prove that some standard FI and BI solution concepts are indeed special cases of our generalized solution concept. In particular, we show that whenever $F_i(h)$ contains all histories – including the past ones – local common strong belief in rationality coincides with the standard common strong belief in rationality. Likewise, we formally prove that the strategies that can be rationally played under common belief in future rationality are exactly those that can be rationally played under local common strong belief in rationality whenever $F_i(h)$ contains only the future histories.

The paper is structured as follows: In Section 2 we present some standard preliminary notions used throughout the paper. In Section 3 we introduce our solution concept as well as our iterative procedure, and we present our main characterization result. Section 4 discusses different special cases of local reasoning. All the proofs are relegated to the Appendices.

2. Preliminaries

2.1. Dynamic games with observable actions

We consider dynamic games with observable actions and simultaneous moves, i.e., dynamic games with the property that at every instance, all players observe the moves that have been undertaken so far. Our results can be extended to arbitrary dynamic games with perfect recall, however. Formally, our framework is described by the following components:

Players. Let I denote the finite set of players, with typical elements i and j. Throughout the paper, we often consider examples with the set of players being $I = \{\text{Ann } (a),\ \text{Bob } (b)\}$.

Histories. For each $i \in I$, let H_i denote the histories where player i moves. We permit more than one player to move at the same history, i.e., $H_i \cap H_j$ may be non-empty. For instance, in Fig. 1 we have $H_a = \{h_0, h_2\}$ and $H_b = \{h_1, h_2\}$, and we write $h_1(b)$ and $h_2(a, b)$ to signify that "only Bob moves at h_1" and that "both Ann and Bob move at h_2" respectively. Let $H := \bigcup_{i \in N} H_i$ be the set of all non-terminal histories, and $H_{-i} := \bigcup_{j \neq i} H_j$ be the set of non-terminal histories where at least one player other than i moves.

Moreover, let $\mathrm{Pr}(h)$ denote the set of histories that weakly precede h, i.e., the past histories as well as h itself. Likewise, let $\mathrm{Fut}(h)$ denote the set of histories that weakly follow h, i.e., the future histories as well as h itself. Then, we also define $\mathrm{Pr}_{-i}(h) := \mathrm{Pr}(h) \cap H_{-i}$ and $\mathrm{Fut}_{-i}(h) := \mathrm{Fut}(h) \cap H_{-i}$. Finally, Z denotes the set of terminal histories, i.e., the histories where no player moves.

Moves and strategies. The finite set of moves (else called actions) from which player i chooses one at some history $h \in H_i$ is denoted by $A_i(h)$. Player i's strategy space is denoted by S_i with typical element s_i, e.g., in Fig. 1 we have $S_a = \{L, RA, RB\}$ and $S_b = \{L, RC, RD\}$. Notice that we define strategies as plans of actions, and not as elements of $\times_{h \in H_i} A_i(h)$. That is, for instance, once Ann has decided to choose L at h_0, she does not need to specify what she would play if h_2 was reached, since she knows that h_2 will not be reached. In either case our analysis would still hold under the alternative definition of a strategy that often appears in the literature (cf., Rubinstein, 1991). As usual, $S := \times_{i \in I} S_i$ denotes the set of strategy profiles with typical element s, and $S_{-i} := \times_{j \neq i} S_j$ denotes the strategy profiles of all players other than i with typical element s_{-i}.

We define player i's set of conditional strategies at some history h as the set of strategies that are consistent with h being reached, and we denote it by $S_i(h)$. Then, $S_{-i}(h)$ denotes the profiles $s_{-i} \in S_{-i}$ that are consistent with h being reached. For each $s_i \in S_i$ we define $H_i(s_i) := \{h \in H_i : s_i \in S_i(h)\}$, and likewise we let $H(s_i) := \{h \in H : s_i \in S_i(h)\}$ and $H_{-i}(s_i) := \{h \in H_{-i} : s_i \in S_i(h)\}$. For instance, in Fig. 1 we have $S_a(h_2) = \{RA, RB\}$ and $H_b(L) = \{h_1\}$. Observe that $S_i(h)$ and $H_i(s_i)$ are always non-empty.

There exists a function $z : S \to Z$, mapping each strategy profile $s \in S$ to a unique terminal history. Finally, each strategy profile induces a path of play, which contains the set of histories that are reached if s is played. Formally, this path contains the non-terminal histories $H(s) := \bigcap_{i \in I} H(s_i)$ and the terminal history $z(s)$.

Utilities. Player i has preferences over the terminal histories, represented by a mapping $v_i : Z \to \mathbb{R}$. Recall that each strategy profile s leads to a unique terminal history $z(s)$. Thus, we obtain the utility function $u_i : S \to \mathbb{R}$, given by $u_i(s) := v_i\big(z(s)\big)$, that represents i's preferences over S.[6] For instance, in Fig. 1 the strategy profile (RA, L) induces the terminal history that yields a utility of 3 to each player.

[6] As usual, we assume that i has vNM preferences over $\Delta(Z)$, and consequently also over $\Delta(S)$. Thus, u_i can be seen as the vNM representation of these preferences.

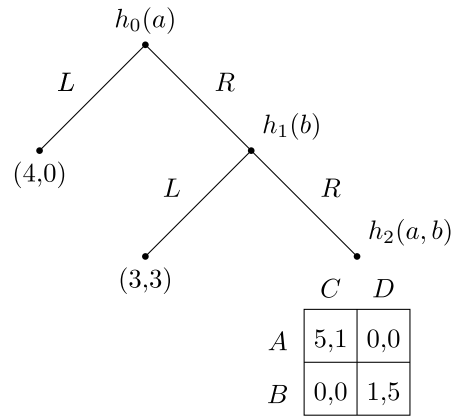


Figure 1: Generalized BoS with outside options.

2.2. Conditional beliefs

Using the standard framework of Battigalli and Siniscalchi (1999, 2002), we model conditional beliefs with an (S-based) type structure, $\big((T_i)_{i\in I}, (\lambda_i)_{i\in I}\big)$, where T_i is player i's set of types and $\lambda_i : T_i \times H_i \to \Delta(S_{-i} \times T_{-i})$ is a Borel function, associating each type $t_i \in T_i$ at each history $h \in H_i$ with a Borel probability measure $\lambda_i^h(t_i) \in \Delta\big(S_{-i}(h) \times T_{-i}\big)$, where $T_{-i} := \bigtimes_{j\neq i} T_j$.[7] Henceforth, we refer to the measure $\lambda_i^h(t_i)$ as t_i's conditional beliefs (or simply beliefs) at a history h. Obviously, this model generalizes the usual type structures à la Harsanyi (1967-68) by replacing simple single-dimensional with a collection of (conditional) beliefs, one for each history.

Example 1. Recall the game in Fig. 1, and consider the S-based type structure $(T_a, T_b, \lambda_a, \lambda_b)$, with the type spaces being $T_a = \{t_a, t'_a\}$ and $T_b = \{t_b, t'_b\}$, and the corresponding conditional beliefs of each type being the ones shown below:

$$\begin{aligned}
\lambda_a^{h_0}(t_a) &= \big(0.40 \otimes (L, t_b)\ ;\ 0.20 \otimes (L, t'_b)\ ;\ 0.10 \otimes (RC, t_b)\ ;\ 0.30 \otimes (RD, t'_b)\big)\\
\lambda_a^{h_2}(t_a) &= \big(0.25 \otimes (RC, t_b)\ ;\ 0.75 \otimes (RD, t'_b)\big)\\
\lambda_a^{h_0}(t'_a) &= \big(0.10 \otimes (L, t_b)\ ;\ 0.40 \otimes (L, t'_b)\ ;\ 0.25 \otimes (RC, t_b)\ ;\ 0.25 \otimes (RC, t'_b)\big)\\
\lambda_a^{h_2}(t'_a) &= \big(1 \otimes (RC, t_b)\big)\\
\lambda_b^{h_1}(t_b) &= \big(0.50 \otimes (RA, t_a)\ ;\ 0.50 \otimes (RA, t'_a)\big)\\
\lambda_b^{h_2}(t_b) &= \big(1 \otimes (RA, t'_a)\big)\\
\lambda_b^{h_1}(t'_b) &= \big(0.25 \otimes (RA, t_a)\ ;\ 0.25 \otimes (RA, t'_a)\ ;\ 0.50 \otimes (RB, t'_a)\big)\\
\lambda_b^{h_2}(t'_b) &= \big(0.50 \otimes (RA, t_a)\ ;\ 0.50 \otimes (RA, t'_a)\big)
\end{aligned}$$

[7] The assumption that "upon reaching a history $h \in H_i$ every type t_i assigns probability 1 to $S_{-i}(h) \times T_{-i}$" corresponds to the standard Condition 1 in (Battigalli and Siniscalchi, 2002, Def. 1). Note that in their paper they further restrict beliefs to satisfy Bayesian updating whenever possible (see Condition 3), thus implicitly assuming that the collection of conditional beliefs forms a conditional probability system, as originally defined by Rênyi (1955).

For instance, if Ann is of type t_a, then at h_0 she puts probability 0.4 to the event that "Bob will play L and is of type t_b". When she finds herself at h_2, she assigns to the same event probability 0. $\triangleleft$

A type structure $\big((T_i)_{i\in I}, (\lambda_i)_{i\in I}\big)$ induces a conditional belief hierarchy for every $t_i \in T_i$. In particular, t_i holds a conditional belief at each $h \in H_i$ about the opponents' strategies (first order conditional beliefs), a conditional belief at each $h \in H_i$ about the opponents' strategies and first order conditional beliefs (second order conditional beliefs), and so on. Throughout the paper, we denote t_i's first order conditional belief at h by

$$b_i^h(t_i) := \operatorname{marg}_{S_{-i}} \lambda_i^h(t_i).$$

In the previous example, t_a's first order conditional beliefs at h_2 put probability 0.25 to RC and probability 0.75 to RD.

A type structure $\big((T_i)_{i\in I}, (\lambda_i)_{i\in I}\big)$ is said to be complete if for every player $i \in I$ the function λ_i is surjective, i.e., for every collection of conditional beliefs $(\mu_i^h)_{h\in H_i}$ there is some type t_i such that $\lambda_i^h(t_i) = \mu_i^h$ for all $h \in H_i$. Battigalli and Siniscalchi (1999) showed the existence of a complete type structure under Bayesian updating. Their result can be easily generalized to type structures without Bayesian updating. Throughout the paper, unless explicitly stated otherwise, we work with complete type structures.[8] Finite type structures that we often consider in our examples can be seen as belief-closed subspaces of a complete type structure.

At some $h \in H_i$ a type t_i of player i is said to believe in some event $E \subseteq S_{-i} \times T_{-i}$ whenever $\lambda_i^h(t_i)(E) = 1$. Then, the types of i that believe in E at h are those in

$$B_i^h(E) := \left\{\, t_i \in T_i : \lambda_i^h(t_i)(E) = 1 \,\right\}.$$

For instance, as we have already mentioned, it is trivially the case that $t_i \in B_i^h\big(S_{-i}(h) \times T_{-i}\big)$ for all $t_i \in T_i$. Moreover, we say that a type believes in E whenever it belongs to

$$B_i(E) := \bigcap_{h\in H_i} B_i^h(E).$$

At some $h \in H_i$ a type t_i of player i is said to strongly believe in some event $E \subseteq S_{-i} \times T_{-i}$ whenever the following condition holds: if E does not contradict history h, then t_i believes in E at h, i.e., formally

$$SB_i^h(E) := \left\{\, t_i \in T_i : \text{ if } (S_{-i}(h) \times T_{-i}) \cap E \neq \emptyset \text{ then } t_i \in B_i^h(E) \,\right\}. \tag{1}$$

[8] Recently, Friedenberg (2010) showed that for standard belief hierarchies a complete type structure that satisfies certain mild topological conditions induces all belief hierarchies, i.e., for every belief hierarchy of each player there exists a type associated with this hierarchy. Moreover, she conjectured – without formally proving it – that the same applies to conditional belief hierarchies that satisfy Bayesian updating. Finally, notice that her result is directly extended to conditional beliefs without Bayesian updating.

Of course, it is straightforward to verify that every t_i strongly believes in $S_{-i}(h) \times T_{-i}$ at $h \in H_i$. Finally, we say that player i strongly believes an event, whenever she strongly believes it at every $h \in H_i$, i.e., formally

$$SB_i(E) := \bigcap_{h \in H_i} SB_i^h(E).$$

2.3. Subjective expected utility and rationality

For an arbitrary conditional belief $\beta_i^h \in \Delta\big(S_{-i}(h)\big)$ and a strategy $s_i \in S_i(h)$, we define i's (subjective) expected utility at $h \in H_i$ in the usual way, i.e., $U_i^h(s_i, \beta_i^h) := \sum_{s_{-i} \in S_{-i}} \beta_i^h(s_{-i}) \cdot u_i(s_i, s_{-i})$. Then, we define the expected utility of a strategy type pair $(s_i, t_i) \in S_i(h) \times T_i$ at a history $h \in H_i$ by

$$U_i^h(s_i, t_i) := U_i^h\big(s_i, b_i^h(t_i)\big). \tag{2}$$

In our Example 1 for instance, Ann's expected utility at h_2 from playing RA, if she is of type t_a, is equal to $U_a^{h_2}(RA, t_a) = 1.25$.

Player's rationality at a history. The event that a player is rational at some history $h \in H_i$ is given by

$$R_i^h := \big\{ \ (s_i, t_i) \in S_i(h) \times T_i : U_i^h(s_i, t_i) \geq U_i^h(s_i', t_i) \text{ for all } s_i' \in S_i(h) \ \big\} \ . \tag{3}$$

If it is indeed the case that $(s_i, t_i) \in R_i^h$, we say that the strategy s_i is optimal/rational given (the first order beliefs induced by) t_i at h. The idea is that, upon reaching a history $h \in H_i$, player i chooses a strategy – among the ones that are still available at h – which maximizes her subjective expected utility. Note that rationality is not an absolute concept. That is, whether a strategy is rational or not depends on the history that we have in mind, as well as on the conditional beliefs held by the player at that history. For instance, in Example 1, we have $R_a^{h_0} = \{(L, t_a), (L, t_a'), (RA, t_a')\}$ and $R_a^{h_2} = \{(RA, t_a), (RA, t_a')\}$. Observe that RA is rational at h_0 given the first order beliefs $b_a^{h_0}(t_a')$, but not given $b_a^{h_0}(t_a)$. Throughout the paper, for notation simplicity we adopt the convention that $R_i^h = S_i(h) \times T_i$ if $h \notin H_i$.

Opponents' rationality at a history. Now, let

$$\begin{aligned} R_{-i}^h \ &:= \ \bigtimes_{j \neq i} \big\{ \ (s_j, t_j) \in S_j(h) \times T_j : \ \text{if } h \in H_j \text{ then } (s_j, t_j) \in R_j^h \ \big\} \\ &= \ \bigtimes_{j \neq i} R_j^h. \end{aligned} \tag{4}$$

denote the event that every player other than i – who is active at h – is rational at h.[9] In other words, R_{-i}^h expresses the idea that upon reaching h, all of i's active opponents at h choose a strategy

[9]In order to obtain $R_{-i}^h = \bigtimes_{j \neq i} R_j^h$ we make use of the convention that $R_j^h = S_j(h) \times T_j$ for all $j \neq$ with $h \notin H_j$.

– among their respective ones – which maximizes their subjective expected utility. In our previous example, for instance, on the one hand we have $R_{-b}^{h_1} = S_a(h_1) \times T_a$, because Ann is not active at h_1, and therefore by our convention $R_a^{h_1} = S_a(h_1) \times T_a$. On the other hand, it is the case that $R_{-b}^{h_2} = R_a^{h_2} = \{(RA, t_a), (RA, t'_a)\}$. This is because Ann is active at h_2, and therefore $R_a^{h_2}$ is given by Eq. (3).

Player's rationality in a set of histories. Now, consider an arbitrary collection $G \subseteq H$ of histories. Then, a strategy-type combination (s_i, t_i) is rational in G whenever it is rational at all histories which (i) are consistent with s_i, and (ii) belong to G. Formally, the event

$$R_i^G := \{ (s_i, t_i) \in S_i \times T_i : (s_i, t_i) \in R_i^h \text{ for all } h \in H_i(s_i) \cap G \} \tag{5}$$

contains the strategy-type pairs that are rational in G. In our previous example, for instance, if we let $G = \{h_2\}$ we get $R_a^{\{h_2\}} = \{(L, t_a), (L, t'_a), (RA, t_a), (RA, t'_a)\}$. Notice that in general $R_i^{\{h\}}$ may differ from R_i^h, e.g., in our working example $R_a^{h_2} = \{(RA, t_a), (RA, t'_a)\} \neq R_a^{\{h_2\}}$. The reason is that, by construction, $R_i^h \subseteq S_i(h) \times T_i$ while $R_i^{\{h\}} \subseteq S_i \times T_i$, i.e., R_i^h considers only strategies that reach h, whereas $R_i^{\{h\}}$ also allows for strategies that are not consistent with h. Because of this, R_i^G does not necessarily coincide with $\bigcap_{h \in G} R_i^h$. Finally, note that the standard notion of rationality corresponds to the event $R_i := R_i^{H_i}$, i.e., a strategy-type combination (s_i, t_i) is rational whenever it is rational all histories $h \in H_i(s_i)$ given the respective conditional first order belief $b_i^h(t_i)$.

Opponents' rationality in a set of histories. Now, let

$$\begin{aligned} R_{-i}^G &:= \bigtimes_{j \neq i} \{ (s_j, t_j) \in S_j \times T_j : (s_j, t_j) \in R_j^h \text{ for all } h \in H_j(s_j) \cap G \} \\ &= \bigtimes_{j \neq i} R_j^G \end{aligned} \tag{6}$$

contain i's opponents' strategy-type combinations that are rational in G. Then, the usual event of every player other than i being rational corresponds to $R_{-i} := R_{-i}^{H_{-i}}$.

3. Local reasoning about the opponents' rationality

In this section, we define a solution concept which incorporates the idea that players reason at each history about their opponents' rationality *at certain histories only.* This is an epistemic concept, implying that it is defined by means of a sequence of restrictions on the players' types, and therefore it gives the set of types (for each player) that are consistent with the particular form of reasoning that we postulate. Then, we provide a simple procedure which yields the strategies that can be rationally played given the types that satisfy the restrictions imposed by the concept.

Note that strictly speaking our concept is a family of concepts, each one corresponding to a different specification of the histories that each players reasons about at each history. In this respect, as we formally show in the next section, several well-known existing solution concepts – such as extensive-form rationalizability or common belief in future rationality, for instance – can be embedded in our framework, i.e., we prove that they correspond to particular specifications of the histories that the players reason about.

Let us begin by defining our notion of "*a player reasoning locally about the opponents' rationality at some histories only*". Formally, fix some $G \subseteq H_{-i}$ and an arbitrary history $h \in H_i$. Then, let

$$SB_i^h\big(R_{-i}^G\big) = \big\{\ t_i \in T_i :\ \text{if } (S_{-i}(h) \times T_{-i}) \cap R_{-i}^G \neq \emptyset \text{ then } t_i \in B_i^h\big(R_{-i}^G\big)\ \big\} \tag{7}$$

denote the event that i strongly believes at h in the event that the opponents are rational at every $h' \in G$. The underlying idea is that, upon finding herself at history h, player i tries to rationalize the opponents' moves at every history in G. If them being rational at every $h' \in G$ does not contradict reaching h, then i will believe at h that they are indeed rational at every history $h' \in G$. This type of local reasoning will be henceforth called *local strong belief in rationality* at h.

So far, we have described how player i reasons – while being at $h \in H_i$ – about the opponents' rationality in an arbitrary $G \subseteq H_{-i}$. Now, we specify the histories that i reasons about while being at each $h \in H_i$. Formally, consider the collection of mappings

$$\mathcal{F} := \big\{\ F_i : H_i \to 2^{H_{-i}}\ \big|\ i \in I\ \big\} \tag{8}$$

with $F_i(h)$ containing the set of histories that player i reasons about while being at $h \in H_i$. For instance, if $F_i(h) = H_{-i}$ for all $h \in H_i$ then player i reasons about all histories, past, present and future ones. On the other hand, if $F_i(h) = \mathrm{Fut}_{-i}(h)$ for all $h \in H_i$ then i reasons only about the present and future histories but not about the past ones. We further discuss special cases of $\mathcal{F}$ structures later in the paper.

Obviously, our local reasoning depends on the choice of $\mathcal{F}$. Thus, we introduce the notion of *local strong belief in rationality with respect to* $\mathcal{F}$ at h, or simply $\mathcal{F}$*-strong belief in rationality* at h. In particular, the types that $\mathcal{F}$-strongly believe at h in the opponents' rationality are those in

$$SB_i^h\Big(R_{-i}^{F_i(h)}\Big) = \big\{\ t_i \in T_i :\ \text{if } (S_{-i}(h) \times T_{-i}) \cap R_{-i}^{F_i(h)} \neq \emptyset \text{ then } t_i \in B_i^h\big(R_{-i}^{F_i(h)}\big)\ \big\}\ . \tag{9}$$

Let us illustrate this notion by means of an example.

Example 1 (cont). Recall the game in Fig. 1 together with the type structure in Ex. 1, and assume that $F_b(h_1) = \{h_0, h_2\}$ and $F_b(h_2) = \{h_2\}$. As we have already noted above, it is the case that $R_{-b}^{F_b(h_1)} = R_a^{\{h_0,h_2\}} = \{(L, t_a), (L, t_a'), (RA, t_a')\}$, thus implying that $SB_b^{h_1}\big(R_{-b}^{F_b(h_1)}\big) = \{t_b\}$. The reason why t_b' does not $\mathcal{F}$-strongly believe at h_1 in Ann's rationality, is that $R_{-b}^{F_b(h_1)}$ is consistent with reaching h_1, and yet it does not receive probability 1 by $\lambda_b^{h_1}(t_b')$. Likewise, it is the case that $R_{-b}^{F_b(h_2)} = R_a^{\{h_2\}} = \{(L,t_a),(L,t_a'),(RA,t_a),(RA,t_a')\}$, and therefore $SB_b^{h_2}\big(R_{-b}^{F_b(h_2)}\big) = \{t_b,t_b'\}$. Indeed, both $\lambda_b^{h_2}(t_b)$ and $\lambda_b^{h_2}(t_b')$ put probability 1 to $R_{-b}^{F_b(h_2)}$. $\triangleleft$

In the previous example, while being at h_1, Bob rules out the possibility of Ann's strategy being RB, because this would contradict her rationality at h_0 which belongs to $F_b(h_1)$. On the other hand, while being at h_2, Bob does not rule out the possibility of Ann having chosen RB, because h_2 is the only history in $F_b(h_2)$, and RB does not contradict Ann's rationality at h_2.

Below, we iterate this idea to obtain our solution concept of $\mathcal{F}$-common strong belief in rationality.

3.1. Local common strong belief in rationality

Take an arbitrary history $h \in H_i$ and an arbitrary collection $\mathcal{F} = \{F_i : H_i \to 2^{H_{-i}} \mid i \in I\}$. Then, we define the following sequences of subsets of T_i:

$$
\begin{aligned}
T_i^{\mathcal{F},1}(h) &:= SB_i^h\Big(R_{-i}^{F_i(h)}\Big)\\
T_i^{\mathcal{F},2}(h) &:= T_i^{\mathcal{F},1}(h)\cap SB_i^h\Big(R_{-i}^{F_i(h)}\cap\big(S_{-i}\times T_{-i}^{\mathcal{F},1}\big(F_i(h)\big)\big)\Big)\\
&\;\;\vdots\\
T_i^{\mathcal{F},k}(h) &:= T_i^{\mathcal{F},k-1}(h)\cap SB_i^h\Big(R_{-i}^{F_i(h)}\cap\big(S_{-i}\times T_{-i}^{\mathcal{F},k-1}\big(F_i(h)\big)\big)\Big)\\
&\;\;\vdots
\end{aligned}
$$

where, for each $k>1$,

$$
\begin{aligned}
T_{-i}^{\mathcal{F},k-1}\big((F_i(h)\big) &:= \bigtimes_{j\neq i}\ \{\ t_j\in T_j : t_j \in T_j^{\mathcal{F},k-1}(h') \text{ for all } h'\in F_i(h)\cap H_j\ \}\\
&= \bigtimes_{j\neq i}\Big(\bigcap_{h'\in F_i(h)\cap H_j} T_j^{\mathcal{F},k-1}(h')\Big).
\end{aligned}
$$

Obviously, $T_i^{\mathcal{F},1}(h)$ contains i's types that strongly believe at h that the opponents are rational at every $h'\in F_i(h)$. Throughout the paper, we refer to the types in $T_i^{\mathcal{F},1}(h)$ as those satisfying *1-fold $\mathcal{F}$-strong belief in rationality at h.*

Now, $T_i^{\mathcal{F},2}(h)$ contains those types in $T_i^{\mathcal{F},1}(h)$ that strongly believe at h that every opponent $j\neq i$ (i) is rational at every $h'\in H_j\cap F_i(h)$, and (ii) strongly believes at every $h'\in H_j\cap F_i(h)$ that every opponent $k\neq j$ is rational at every $h''\in H_k\cap F_j(h')$. The event described in (i) corresponds to $R_{-i}^{F_i(h)}$, while the event described in (ii) corresponds to

$$
S_{-i}\times T_{-i}^{\mathcal{F},1}\big(F_i(h)\big) = \bigtimes_{j\neq i}\ \{\ (s_j,t_j)\in S_j\times T_j : t_j\in SB_j^h\big(R_{-j}^{F_j(h')}\big) \text{ for all } h'\in F_i(h)\cap H_j\ \}
$$

in the second equation of the sequence above. The reason for explicitly requiring every type in $T_i^{\mathcal{F},2}(h)$ to belong to $T_i^{\mathcal{F},1}(h)$ is that the strong belief operator is not monotonic, thus implying that $SB_i^h(E \cap F)$ is not necessarily equal to $SB_i^h(E) \cap SB_i^h(F)$.[10] Therefore $SB_i^h\big(R_{-i}^{F_i(h)}\big)$ does not follow directly from $SB_i^h\big(R_{-i}^{F_i(h)} \cap \big(S_{-i} \times T_{-i}^{\mathcal{F},1}\big(F_i(h)\big)\big)\big)$. Throughout the paper, we refer to the types in $T_i^{\mathcal{F},2}(h)$ as those satisfying *up to 2-fold $\mathcal{F}$-strong belief in rationality at h.* The reason we add the term "up to" is that, by construction, $T_i^{\mathcal{F},2}(h) \subseteq T_i^{\mathcal{F},1}(h)$, as we have already discussed above.

Continuing inductively we define the set of types that satisfy *up to k-fold $\mathcal{F}$-strong belief in rationality at h.* Those are the types in $T_i^{\mathcal{F},k}(h)$. Then, the types that satisfy *$\mathcal{F}$-common strong belief in rationality at h* are those in

$$T_i^{\mathcal{F}}(h) := \bigcap_{k=1}^{\infty} T_i^{\mathcal{F},k}(h). \tag{10}$$

The types that satisfy *$\mathcal{F}$-common strong belief in rationality ($\mathcal{F}$-CSBR)* are those in

$$T_i^{\mathcal{F}} := \bigcap_{h \in H_i} T_i^{\mathcal{F}}(h). \tag{11}$$

Observe that in order to obtain the types that satisfy $\mathcal{F}$-CSBR, we need to take two intersections. In particular, first we find, for each $h \in H_i$, the types that that satisfy the (infinitely many) restrictions that $\mathcal{F}$-CSBR imposes at h (see Eq. (10)), and then we select those types that satisfy all these restrictions at every $h \in H_i$ (see Eq. (11)). The reason for doing so is that $T_{-i}^{\mathcal{F},k-1}\big((F_i(h)\big)$ differs for every $h \in H_i$. This is because player i may reason about different histories at each $h \in H_i$, i.e., it may be the case that $F_i(h) \neq F_i(h')$ for two histories $h, h' \in H_i$. This is in contrast to the usual (global) definition of common strong belief in rationality à la Battigalli and Siniscalchi (2002).[11]

Finally, we say that a strategy $s_i \in S_i$ can be *rationally played under $\mathcal{F}$-common strong belief in rationality ($\mathcal{F}$-RCSBR)* whenever $s_i \in \mathrm{Proj}_{S_i}\big(R_i \cap (S_i \times T_i^{\mathcal{F}})\big)$.

3.2. Local iterated conditional dominance procedure

In this section we introduce a (finite) procedure which, for every player $i \in I$ and every history $h \in H_i$, iteratively eliminates (at each round), strategies from $S_i(h)$ and first order conditional beliefs from $\Delta\big(S_{-i}(h)\big)$. Formally, it is a simultaneous generalization of the iterated conditional dominance procedure (ICDP), originally introduced by Shimoji and Watson (1998), and the backward dominance procedure, originally defined Perea (2014).[12] Before, formally defining our procedure, let

[10] It is well known that the conjunction property implies monotonicity. Therefore, violations of monotonicity – which the strong belief operator exhibits – lead to violations of the conjunction property. We refer to Battigalli and Siniscalchi (2002) for a detailed discussion on this issue.

[11] Later in the paper, we explicitly discuss the relationship between $\mathcal{F}$-CSBR and CSBR, as defined by Battigalli and Siniscalchi (2002).

[12] Later in the paper, we discuss the relationship of our procedure with ICDP.

us first introduce the notion of a decision problem, which will play a central role throughout this section. In particular, our procedure will be defined as a sequence of decision problems for each player $i \in I$ and each history $h \in H_i$.

Decision problem. A decision problem for player $i \in I$ at a history $h \in H_i$ is a tuple $\big(B_i(h), D_i(h)\big)$, with $B_i(h) \subseteq S_{-i}(h)$ and $D_i(h) \subseteq S_i(h)$. Intuitively, $B_i(h)$ can be seen as the subset of the opponents' strategies that i could deem possible at h. At this point, we should already make clear that the link between $B_i(h)$ and what i could deem possible at h is only an informal one. The actual relationship between the two will become apparent later on in the paper. Thus, for the time being, $B_i(h)$ and $D_i(h)$ will be merely treated as auxiliary tools, without a concrete meaning.

A strategy $s_i \in D_i(h)$ is said to be rational in the decision problem $\big(B_i(h), D_i(h)\big)$ whenever there exists a probability measure $\beta_i^h \in \Delta\big(B_i(h)\big)$ such that $U_i^h(s_i, \beta_i^h) \geq U_i^h(s_i', \beta_i^h)$ for all $s_i' \in D_i(h)$. Thus, we draw a link between two different notions of rationality, i.e., between rationality of a strategy-type combination in a complete type structure on the one hand, and rationality of a strategy in a decision problem on the other hand.

Now, for an arbitrary collection $\mathcal{F}$, our procedure will be defined by means of a (weakly) decreasing sequence $\big(B_i^{\mathcal{F},k}(h), D_i^{\mathcal{F},k}(h)\big)_{k\geq 0}$ of decision problems for each $i \in I$ and each $h \in H_i$. That is, at each step of our procedure, we will simultaneously eliminate strategies from $S_i(h)$ and strategy combinations from $S_{-i}(h)$.

Initial step of the procedure. For $k = 0$, we define

$$\begin{aligned} B_i^{\mathcal{F},0}(h) &:= S_{-i}(h) \\ D_i^{\mathcal{F},0}(h) &:= S_i(h). \end{aligned}$$

Obviously, this initial step does not depend on $\mathcal{F}$.

Inductive step of the procedure. Now, fix some $k > 0$ and suppose that for each $j \in I$ and each $h' \in H_j$ we have undertaken the $(k-1)$-th step of our procedure, thus having obtained $\big(B_j^{\mathcal{F},k-1}(h'), D_j^{\mathcal{F},k-1}(h')\big)$. Then, for an arbitrary $h \in H_i$, define $\big(B_j^{\mathcal{F},k}(h), D_j^{\mathcal{F},k}(h)\big)$ by

$$B_i^{\mathcal{F},k}(h) := \begin{cases} C_i^{\mathcal{F},k-1}(h) & \text{if } C_i^{\mathcal{F},k-1}(h) \neq \emptyset \\ B_i^{\mathcal{F},k-1}(h) & \text{if } C_i^{\mathcal{F},k-1}(h) = \emptyset \end{cases} \tag{12}$$

$$D_i^{\mathcal{F},k}(h) := \left\{ s_i \in D_i^{\mathcal{F},k-1}(h) : s_i \text{ is rational in } \big(B_i^{\mathcal{F},k}(h), D_i^{\mathcal{F},k-1}(h)\big) \right\}, \tag{13}$$

where

$$C_i^{\mathcal{F},k-1}(h) := \bigtimes_{j\neq i} \left\{ s_j \in S_j(h) : s_j \in D_j^{\mathcal{F},k-1}(h') \text{ for all } h' \in H_j(s_j) \cap F_i(h) \right\}. \tag{14}$$

The underlying idea behind our procedure is as follows: First, for each $h \in H_i$, we compute $C_i^{\mathcal{F},k-1}(h)$ which contains all strategy combinations of i's opponents which (i) are consistent with reaching h, and (ii) have not been eliminated from $D_j^{\mathcal{F},k-1}(h')$ at any $h' \in F_i(h)$ and for any player j who is active at h'. Notice that in principle $C_i^{\mathcal{F},k-1}(h)$ might be empty. To see this, consider for instance the game in Fig. 1 with $F_b(h_1) = \{h_0, h_2\}$, and assume that $D_a^{\mathcal{F},k-1}(h_0) = \{L\}$. Then, clearly it is the case that $S_a(h_1) \cap D_a^{\mathcal{F},k-1}(h_0) = \emptyset$, thus implying that $C_b^{\mathcal{F},k-1}(h_1) = \emptyset$.

Having defined $C_i^{\mathcal{F},k-1}(h)$, we can now proceed to the k-th step of our procedure, by first defining $B_i^{\mathcal{F},k}(h)$. In particular, a strategy combination $s_{-i} = (s_j)_{j \neq i}$ is eliminated from $B_i^{\mathcal{F},k-1}(h)$ if and only if (i) there exists some history $h' \in F_i(h) \cap H_j(s_j)$ such that s_j has been eliminated from $D_j^{\mathcal{F},k-1}(h')$, and also (ii) there exists another strategy $s'_{-i} = (s'_j)_{j \neq i} \in B_i^{\mathcal{F},k-1}(h)$ such that for every $h' \in F_i(h) \cap H_j(s_j)$ it is the case that $s'_j \in D_j^{\mathcal{F},k-1}(h')$, i.e., not all strategy combinations are eliminated from $B_i^{\mathcal{F},k-1}(h)$.

Now, once we have obtained $B_i^{\mathcal{F},k-1}(h)$, we can define the decision problem $\big(B_i^{\mathcal{F},k}(h), D_i^{\mathcal{F},k-1}(h)\big)$, and we eliminate from $D_i^{\mathcal{F},k}(h)$ the strategies that are not rational in this decision problem. As we have already mentioned above, it follows from Pearce (1984, Lem. 3) that a strategy is eliminated from $D_i^{\mathcal{F},k-1}(h)$ if and only if it is strictly dominated within this decision problem.

This elimination procedure is called *local iterated conditional dominance procedure with respect to $\mathcal{F}$*, or simply *$\mathcal{F}$-iterated conditional dominance procedure ($\mathcal{F}$-ICDP)*. Obviously, since we consider only finite dynamic games, $\mathcal{F}$-ICDP converges after finitely many steps. That it, there exists some $K \geq 0$ such that for each $k \geq K$, for every $i \in I$ and every $h \in H_i$, it is the case that $\big(B_i^{\mathcal{F},k}(h), D_i^{\mathcal{F},k}(h)\big) = \big(B_i^{\mathcal{F},K}(h), D_i^{\mathcal{F},K}(h)\big)$. Then, we write $\big(B_i^{\mathcal{F}}(h), D_i^{\mathcal{F}}(h)\big) = \big(B_i^{\mathcal{F},K}(h), D_i^{\mathcal{F},K}(h)\big)$. We say that a *strategy s_i survives the $\mathcal{F}$-iterated conditional dominance procedure* if it is the case that $s_i \in D_i^{\mathcal{F}}(h)$ for all $h \in H_i(s_i)$.

Below, we illustrate the $\mathcal{F}$-ICDP with an example.

Example 2. Recall the example of Fig. 1, and assume that $\mathcal{F} = \{F_a, F_b\}$ is such that Ann reasons about all histories, whereas Bob reasons only about present and future histories, i.e., formally, $F_a(h_0) = F_a(h_2) = \{h_1, h_2\}$ and $F_b(h_1) = F_b(h_2) = \{h_2\}$. Let us now depict each decision problem $\big(B_i^{\mathcal{F},k}(h), D_i^{\mathcal{F},k}(h)\big)$ with a normal form game. The steps of the $\mathcal{F}$-ICDP are represented by the lines that cross out the corresponding strategies. Eliminations from $B_i^{\mathcal{F},k}(h)$ are represented by dashed lines, whereas eliminations from $D_i^{\mathcal{F},k}(h)$ are represented by continuous lines. The corresponding number next to each line refers to the step during which the respective strategy was eliminated.

In particular, at the first step ($k = 1$), no strategy is eliminated from $B_i^{\mathcal{F},1}(h)$ for any $i \in I$ and any $h \in H_i$. Then, RB is eliminated from $D_a^{\mathcal{F},0}(h_0)$ because it is strictly dominated by L at h_0, and likewise RC eliminated from $D_b^{\mathcal{F},0}(h_1)$ because it is strictly dominated by L at h_1. Hence, we obtain $D_a^{\mathcal{F},1}(h_0) = \{L, RA\}$ and $D_b^{\mathcal{F},1}(h_1) = \{L, RD\}$. Furthermore, no strategy is eliminated at h_2, i.e., it

Ann

$B_a^{\mathcal{F},k}(h_0)$ over the columns; $D_a^{\mathcal{F},k}(h_0)$ beside the rows

h_0	~~L~~	~~RC~~	RD	
L	4,0	4,0	4,0	
~~RA~~	~~3,3~~	~~5,1~~	~~0,0~~	$(k=2)$
~~RB~~	~~3,3~~	~~0,0~~	~~1,5~~	$(k=1)$
	$(k=4)$	$(k=2)$		

$B_a^{\mathcal{F},k}(h_2)$ over the columns; $D_a^{\mathcal{F},k}(h_2)$ beside the rows

h_2	~~RC~~	RD	
~~RA~~	~~5,1~~	~~0,0~~	$(k=2)$
RB	0,0	1,5	
	$(k=2)$		

Bob

$D_b^{\mathcal{F},k}(h_1)$ over the columns; $B_b^{\mathcal{F},k}(h_1)$ beside the rows

h_1	~~L~~	~~RC~~	RD	
~~RA~~	~~3,3~~	~~5,1~~	~~0,0~~	$(k=3)$
RB	3,3	0,0	1,5	
	$(k=3)$	$(k=1)$		

$D_b^{\mathcal{F},k}(h_2)$ over the columns; $B_b^{\mathcal{F},k}(h_2)$ beside the rows

h_2	~~RC~~	RD	
~~RA~~	~~5,1~~	~~0,0~~	$(k=3)$
RB	0,0	1,5	
	$(k=3)$		

is the case that $D_a^{\mathcal{F},1}(h_2) = \{RA, RB\}$ and $D_b^{\mathcal{F},1}(h_2) = \{RC, RD\}$.

At the second step ($k = 2$), Bob's strategy RC is eliminated both from $B_a^{\mathcal{F},1}(h_0)$ and from $B_a^{\mathcal{F},1}(h_2)$. This is because h_1 belongs to both $F_a(h_0)$ and $F_a(h_2)$. Thus, we obtain $B_a^{\mathcal{F},2}(h_0) = \{L, RD\}$ and $B_a^{\mathcal{F},2}(h_2) = \{RD\}$. On the other hand, RB is not eliminated from either $B_b^{\mathcal{F},1}(h_1)$ or $B_b^{\mathcal{F},1}(h_2)$, because h_0 does not belong to $F_b(h_1)$ or to $F_b(h_2)$. Then, RA is eliminated from $D_a^{\mathcal{F},1}(h_0)$ because it is strictly dominated by L at h_0, and likewise it is also eliminated from $D_a^{\mathcal{F},1}(h_2)$ because it is strictly dominated by RB at h_2.

Similarly we continue until the fourth step when the procedure stops. The only strategy profile that survives $\mathcal{F}$-ICDP is (L, RD). Indeed, observe that $L \in D_a^{\mathcal{F}}(h_0)$, where $\{h_0\} = H_a(L)$. Likewise, observe that $RD \in D_b^{\mathcal{F}}(h_1) \cap D_b^{\mathcal{F}}(h_2)$, where $\{h_1, h_2\} = H_b(RD)$.

At this point, we should also point out that the procedure yields, not only the strategy profiles that survive, but also the conditional beliefs at each history, e.g., according to the procedure, the only belief that Bob can have at h_1 is to put probability 1 to Ann playing according to the strategy RB. Below, we further elaborate on the fact that the procedure simultaneously induces strategies and conditional beliefs for each player at each history. $\triangleleft$

Interpretation. Let us begin by stressing that at each step of our procedure we perform two types of elimination, viz., for each player $i \in I$ and each $h \in H_i$, first we eliminate opponents' strategy combinations from $B_i^{\mathcal{F},k}(h)$, and then we eliminate strategies from $D_i^{\mathcal{F},k}(h)$. Note that these two types of elimination are conceptually very different. Let us for the time being focus on $B_i^{\mathcal{F},k}(h)$.

Eliminating a strategy combination $s_{-i} \in S_{-i}(h)$ from $B_i^{\mathcal{F},k}(h)$ can be thought as eliminating all of i's first order conditional beliefs at h that put positive probability to s_{-i}. Consequently, this elimination can be thought as a restriction imposed on i's types, viz., eliminating s_{-i} from $B_i^{\mathcal{F},k}(h)$ essentially means that we are ruling out all types t_i with the property that $\text{marg}_{S_{-i}} \lambda_i^h(t_i)(\{s_{-i}\}) > 0$. But then recall that this is exactly what $\mathcal{F}$-CSBR does, i.e., it recursively imposes restrictions on i's types. In the next section we show that there is indeed a very tight relationship between eliminating opponents' strategies from $B_i^{\mathcal{F},k}(h)$ and eliminating own types from $T_i^{\mathcal{F},k-1}(h)$. Thus, it becomes clear why earlier in this section we stated that the strategy combinations in $B_i^{\mathcal{F},k}(h)$ can be thought as those that i could deem possible at h after k rounds of reasoning.

3.3. Characterization results

As we have already mentioned in the previous section, there is a very tight relationship between the process of eliminating own types from $T_i^{\mathcal{F},k-1}(h)$ and the process of eliminating opponents' strategy profiles from $B_i^{\mathcal{F},k}(h)$. The following result makes this relationship formal.

Theorem 1. *Consider a complete type structure $\big((T_i)_{i\in I}, (\lambda_i)_{i\in I}\big)$ and fix an arbitrary $\mathcal{F}$. Then, for every player $i \in I$, every history $h \in H_i$ and every $k > 0$, the following hold:*

(i) If $t_i \in T_i^{\mathcal{F},k-1}(h)$ then there exists some $\beta_i^h \in \Delta\big(B_i^{\mathcal{F},k}(h)\big)$ with $b_i^h(t_i) = \beta_i^h$.

(ii) If $\beta_i^h \in \Delta\big(B_i^{\mathcal{F},k}(h)\big)$ then there exists some $t_i \in T_i^{\mathcal{F},k-1}(h)$ with $b_i^h(t_i) = \beta_i^h$.

For instance, in the context of Ex. 2 the previous result implies that, for every type $t_a \in T_a^{\mathcal{F},1}(h_0)$ it is the case that $b_a^{h_0}(t_a)$ puts probability 0 to RC. This is because $B_a^{\mathcal{F},2}(h_0) = \{L, RD\}$. Still, we should stress that part (ii) in the theorem above does not say that every t_i with $b_i^h(t_i) \in \Delta\big(B_i^{\mathcal{F},k}(h)\big)$ belongs to $T_i^{\mathcal{F},k-1}(h)$. To see this, consider a type t_i which at h puts probability 1 to a strategy-type combination $(s_{-i}, t_{-i}) \in B_i^{\mathcal{F},k}(h) \times \big(T_{-i}^{\mathcal{F},k-2}(h') \setminus T_{-i}^{\mathcal{F},k-1}(h')\big)$ where $h' \in F_i(h)$, implying that $b_i^h(t_i) \in \Delta\big(B_i^{\mathcal{F},k}(h)\big)$ and also $t_i \notin T_i^{\mathcal{F},k-1}(h)$. Notice that such a type exists whenever the type structure is complete.

Then, it is rather straightforward to characterize the strategies that can be rationally played under $\mathcal{F}$-common strong belief in rationality, by means of the $\mathcal{F}$-iterated conditional dominance procedure.

Theorem 2. *Consider a complete type structure $\big((T_i)_{i\in I}, (\lambda_i)_{i\in I}\big)$ and fix an arbitrary $\mathcal{F}$. Then, for an arbitrary player $i \in I$, it is the case that $s_i \in \text{Proj}_{S_i}\big(R_i \cap (S_i \times T_i^{\mathcal{F}})\big)$ if and only if $s_i \in D_i^{\mathcal{F}}(h)$ for all $h \in H_i(s_i)$.*

The previous result formally states that a strategy can be rationally played under $\mathcal{F}$-CSBR if and only it survives the $\mathcal{F}$-ICDP. For instance, in the context of Ex. 2, the only strategy profile that can be rationally played under $\mathcal{F}$-CSBR is (L, RD), as this is the only strategy profile surviving the $\mathcal{F}$-ICDP.

4. Special cases of local reasoning

In this section we present some special cases of $\mathcal{F}$. As we have already mentioned earlier in the paper, in some of these cases, $\mathcal{F}$-CSBR coincides with existing solution concepts, such as common strong belief in rationality (Battigalli and Siniscalchi, 2002) or common belief in future rationality (Perea, 2014). Yet, note that our framework is flexible enough to accommodate any $\mathcal{F}$.

4.1. Reasoning about all histories: Forward induction

The general idea behind forward induction reasoning is that players observe their opponents' past behavior and use this information in order to form beliefs about their opponents' future behavior.[13]

The most prominent forward induction solution concept is extensive-form rationalizability (EFR), originally introduced by Pearce (1984), subsequently simplified by Battigalli (1997) and later epistemically characterized by Battigalli and Siniscalchi (2002) by means of rationality and common strong belief in rationality (in a complete type structure). The main idea is that players try to rationalize the opponents' strategies whenever this is possible. That is, upon reaching an arbitrary $h \in H_i$, player i is assumed to believe that *her opponents are rational at all histories*, as long as their rationality is not contradicted by the fact that history h has been reached. Thus, EFR implicitly postulates that player i at h reasons about *the opponents' rationality at all histories.*

Let us first formally recall the concept of up to k-fold strong belief in rationality, as it was originally defined by Battigalli and Siniscalchi (2002). Consider the following sequences of subsets of T_i:

$$\begin{aligned}
SB_i^1 &:= SB_i(R_{-i}) \\
SB_i^2 &:= SB_i^1 \cap SB_i\big(R_{-i} \cap (S_{-i} \times SB_{-i}^1)\big) \\
&\vdots \\
SB_i^k &:= SB_i^{k-1} \cap SB_i\big(R_{-i} \cap (S_{-i} \times SB_{-i}^{k-1})\big) \\
&\vdots
\end{aligned}$$

[13]FI is not a solution concept. Rather it is a general principle which is present in different concepts that have appeared in the literature (e.g., Pearce, 1984; Battigalli and Siniscalchi, 2002; Stalnaker, 1998; Battigalli and Friedenberg, 2012; Govindan and Wilson, 2009; Cho, 1987; Cho and Kreps, 1987; McLennan, 1985; Hillas, 1994).

with $SB_{-i}^{k-1} := \bigtimes_{j \neq i} SB_j^{k-1}$ for each $k > 1$. Moreover, let

$$CSB_i := \bigcap_{k=1}^{\infty} SB_i^k \tag{15}$$

be the set of types that satisfy *common strong belief in rationality (CSBR)*. Finally, we say that a strategy s_i can be rationally played under CSBR whenever $s_i \in \mathrm{Proj}_{S_i}\big(R_i \cap (S_i \times CSB_i)\big)$.

Let us now assume that $F_i(h) = H_{-i}$ for all $h \in h_i$ and all $i \in I$. Then, we ask whether there is a formal relationship between common strong belief in rationality on the one hand, and our $\mathcal{F}$-common strong belief in rationality on the other. As it turns out the two notions are equivalent, as shown below.

Proposition 1. *Consider an arbitrary type structure $\big((T_i)_{i \in I}, (\lambda_i)_{i \in I}\big)$. Moreover, let $\mathcal{F}$ be such that $F_i(h) = H_{-i}$ for all $i \in I$ and all $h \in H_i$. Then, for every player $i \in I$ and every $k > 0$, it is the case that $SB_i^k = \bigcap_{h \in H_i} T_i^{\mathcal{F},k}(h)$.*

Two immediate conclusions follow directly from the previous result. First, a type satisfies common strong belief in rationality if and only it satisfies $\mathcal{F}$-common strong belief in rationality, i.e., $CSB_i = T_i^{\mathcal{F}}$. Second, a strategy can be rationally played under common strong belief in rationality if and only if it can be rationally played under $\mathcal{F}$-common strong belief in rationality. This is formally stated in the following corollary. The proof trivially follows from the definition of rationality.

Corollary 1. *Consider an arbitrary type structure $\big((T_i)_{i \in I}, (\lambda_i)_{i \in I}\big)$. Moreover, let $\mathcal{F}$ be such that $F_i(h) = H_{-i}$ for all $i \in I$ and all $h \in H_i$. Then, for every player $i \in I$, it is the case that $\mathrm{Proj}_{S_i}\big(R_i \cap (S_i \times CSB_i)\big) = \mathrm{Proj}_{S_i}\big(R_i \cap (S_i \times T_i^{\mathcal{F}})\big)$.*

Another direct consequence of the previous result – combined with Theorem 2 of the previous section and the characterization result of Shimoji and Watson (1998) – is that, in a complete type structure, a strategy survives k steps of our $\mathcal{F}$-iterated conditional dominance procedure if and only if it survives k steps of Shimoji and Watson's iterated conditional dominance procedure (ICDP). In this sense, ICDP is a special case of $\mathcal{F}$-ICDP.

Now, notice that in Proposition 1 we do not impose any restriction on the type structure, and in particular we do not focus exclusively on complete type structures. In fact, it is known that whenever we restrict attention to complete type structures, Rationality and CSBR epistemically characterize the strategies that are predicted by Extensive Form Rationalizability (EFR) (Pearce, 1984). On the other hand, if we allow for an arbitrary type structure, Rationality and CSBR yields an Extensive Form Best Response Set (EFBRS) (Battigalli and Friedenberg, 2012). The fact that Proposition 1 does not restrict the type structure implies that Rationality and $\mathcal{F}$-CSBR also yield an EFBRS.

4.2. Reasoning about future histories: Backward induction

Contrary to forward induction, the general idea behind backward induction reasoning is that players ignore their opponents' realized past behavior when they form beliefs about their opponents' future behavior.[14]

The two concepts that in our view capture this idea – and nothing more – for arbitrary dynamic games are the backward dominance procedure (BDP) (Perea, 2014) and backward rationalizability (BR) (Penta, 2009).[15] Note that these two concepts differ only in that BR postulates Bayesian updating, whereas BDP does not. Both these two concepts are epistemically characterized by rationality and common belief in future rationality in a complete type structure (Perea, 2014).[16] Throughout the paper, we will mostly focus our discussion on BDP. Nonetheless, our analysis is also valid in the case of BR.

The idea behind BDP is that players maintain the belief that their opponents will continue being rational irrespective of the moves they have observed so far. That is, upon reaching a history $h \in H_i$ player i is assumed to believe that her opponents will behave rationally from that point onwards, even if reaching this history contradicts the opponents' rationality. Thus, BDP implicitly postulates that player i at h reasons only about the opponents' rationality *at the current history and in the future*.

First, we define the event that player i believes in the opponents' future rationality by

$$FB_i(R_{-i}) := \bigcap_{h \in H_i} B_i^h\big(R_{-i}^{\mathrm{Fut}(h)}\big). \tag{16}$$

Then, we consider the following sequence of subsets of T_i:

$$\begin{aligned}
FB_i^1 &:= FB_i(R_{-i}) \\
FB_i^2 &:= FB_i^1 \cap B_i(S_{-i} \times FB_{-i}^1) \\
&\vdots \\
FB_i^k &:= FB_i^{k-1} \cap B_i(S_{-i} \times FB_{-i}^{k-1}) \\
&\vdots
\end{aligned}$$

[14] Once again, BI is not a solution concept but rather a general principle embodied in different concepts in the literature (e.g., Selten, 1965; Kreps and Wilson, 1982; Perea, 2014; Baltag et al., 2009; Penta, 2009).

[15] Concepts like subgame perfect equilibrium (Selten, 1965) or sequential equilibrium (Kreps and Wilson, 1982) impose additional equilibrium conditions, whereas the standard backward induction procedure is well-defined only for perfect information extensive-form games without relevant ties.

[16] Formally speaking, Perea (2014) does not fix a type structure. Instead he looks across different (finite) type structures. This approach is essentially equivalent to ours, as every finite type structure can be embedded into the complete type structure that we use here via a type morphism that preserves the conditional belief hierarchies.

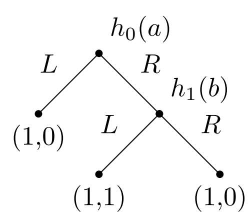
where $FB_{-i}^{k-1} := \times_{j \neq i} FB_j^{k-1}$ for each $k > 1$. We say that FB_i^k contains the types that satisfy *up to k-fold belief in future rationality.*[17] Now, let

$$CFB_i := \bigcap_{k=1}^{\infty} FB_i^k \tag{17}$$

contain the types that satisfy *common belief in future rationality (CBFR).* We say that a strategy s_i can be rationally played under CBFR whenever $s_i \in \mathrm{Proj}_{S_i}\big(R_i \cap (S_i \times CFB_i)\big)$.

The previous idea is formally captured by the assumption that $F_i(h) = \mathrm{Fut}_{-i}(h)$ for all $h \in H_i$ and all $i \in I$. Then, it is natural to investigate the formal relationship between belief in future rationality on the one hand and $\mathcal{F}$-strong belief in rationality on the other. First, let us point out that whenever $\mathcal{F}$ is such that $F_i(h) = \mathrm{Fut}_{-i}(h)$, it is by definition the case that $\mathcal{F}$-strong belief is directly reduced to standard belief. Thus, it is not surprising that the two notions are equivalent in terms of the strategy profiles the predict. Still, this is not necessarily the case for the types they induce. Let us first illustrate with an example a case where CBFR does not coincide with $\mathcal{F}$-CSBR.

Example 3. Consider the following dynamic game between Ann and Bob.



Now, consider the type structure $(T_a, T_b, \lambda_a, \lambda_b)$ with the type spaces being $T_a = \{t_a, t'_a\}$ and $T_b = \{t_b\}$ and the corresponding conditional beliefs being given by

$$\begin{aligned}
\lambda_a^{h_0}(t_a) &= \big(1 \otimes (R, t_b)\big) \\
\lambda_a^{h_0}(t'_a) &= \big(1 \otimes (L, t_b)\big) \\
\lambda_b^{h_1}(t_b) &= \big(1 \otimes (R, t_a)\big)
\end{aligned}$$

First notice that the only type of Ann that is consistent with up to 1-fold belief in future rationality is t'_a, viz., formally, $FR_a^1 = \{t'_a\}$. This is because, Bob's unique rational strategy at h_1 is to choose L. This implies that t_b does not satisfy up to 2-fold belief in future rationality. Indeed, observe that $\lambda_b^{h_1}(t_b)(S_a \times FR_a^1) = \lambda_b^{h_1}(t_b)(S_a \times \{t'_a\}) = 0$. In fact, it is the case that $FR_b^2 = \emptyset$, i.e., there is no type of Bob satisfying up to 2-fold belief in future rationality.

[17]Notice that in the previous definition we have the set $B_i(S_{-i} \times FB_{-i}^{k-1})$ rather than $B_i\big(R_{-i} \cap (S_{-i} \times FB_{-i}^{k-1}\big)$. This is in contrast to the respective definition of "up to k-fold strong belief in rationality". This is because – unlike strong belief – B_i is a monotonic operator (Battigalli and Siniscalchi, 2002).

Now, suppose that $F_i(h) = \mathrm{Fut}_{-i}(h)$ for each $h \in H_i$ and each $i \in I$, i.e., $F_a(h_0) = \{h_1\}$ and $F_b(h_1) = \emptyset$. Then, observe that again the only type of Ann satisfying 1-fold $\mathcal{F}$-strong belief in rationality at h_0 is t'_a, viz., $T_a^{\mathcal{F},1}(h_0) = \{t'_a\}$. But, then t_b does satisfy up to 2-fold $\mathcal{F}$-strong belief in rationality at h_1. This is because $F_b(h_1) = \emptyset$, and therefore $T_b^{\mathcal{F},k}(h_1) = T_b$ for all $k > 0$. $\triangleleft$

The reason for the previously illustrated divergence between FB_b^2 and $\bigcap_{h \in H_b} T_b^{\mathcal{F},2}(h)$ is that in order for a type t_b to satisfy up to 2-fold belief in future rationality, it must attach at h_1 probability 1 to $S_a \times FB_a^1$. But then, FB_a^1 contains Ann's types that require Ann to believe at h_0 that Bob will be rational from that point onwards. In other words, t_b must believe at h_1 that Ann believed at the *earlier history* h_0 that Bob would be rational at all histories following h_0. On the other hand, in order for a type t_b to satisfy up to 2-fold $\mathcal{F}$-strong belief in rationality, it must believe at h_1 that Ann will believe at all histories *following* h_1 that Bob will be rational at all future histories. However, in the previous example there is no history following h_1, and hence no requirement is being imposed. In this respect our concept of $\mathcal{F}$-CSBR with $F_i(h) = \mathrm{Fut}_{-i}(h)$ is a truly backward induction concept as *it completely disregards the past.* In particular, it postulates that players ignore not only the opponents' past behavior, but also the opponents' reasoning at past histories.

Still, even though $\mathcal{F}$-CSBR and $CBFR$ differ in the conditional beliefs that they induce, they coincide in the predictions they make. In particular, as we show below, given a complete type structure, a strategy can be rationally played under $\mathcal{F}$-CSBR if and only if it can be rationally played under $CBFR$.

Proposition 2. *Consider a complete type structure* $\big((T_i)_{i\in I}, (\lambda_i)_{i\in I}\big)$. *Moreover, let* $\mathcal{F}$ *be such that* $F_i(h) = \mathrm{Fut}_{-i}(h)$ *for all* $i \in I$ *and all* $h \in H_i$. *Then, for every player* $i \in I$, *it is the case that* $\mathrm{Proj}_{S_i}\big(R_i \cap (S_i \times CFB_i)\big) = \mathrm{Proj}_{S_i}\big(R_i \cap (S_i \times T_i^{\mathcal{F}})\big)$.

The proof of the result follows almost directly from Lemma B1 in Appendix B, which formally proves that BDP and $\mathcal{F}$-ICDP are essentially equivalent.

Finally, notice that while $\mathcal{F}$-CSBR and $CBFR$ yield the same predicted strategies in a complete type structure, this is not necessarily the case for an arbitrary type structure. To see this recall Example 3. In particular, observe that, given the type structure that we assume, Rationality and CBFR yields an empty set of predictions, whereas Rationality and $\mathcal{F}$-CSBR induces a non-empty prediction, viz., $\mathrm{Proj}_{S_b}\big(R_b \cap (S_b \times T_b^{\mathcal{F}})\big) = \{L\}$, while $\mathrm{Proj}_{S_b}\big(R_b \cap (S_b \times CFB_b)\big) = \emptyset$.

A. Proofs of Section 3

We first introduce some additional notation and prove some intermediate results that we will use throughout the proof of our main theorem.

Lemma A1 (Optimality principle)**.** *Fix an arbitrary player $i \in I$, an arbitrary history $h \in H_i$, an arbitrary $\mathcal{F}$ and some $k > 0$. Then, a strategy $s_i \in S_i(h)$ is rational in $\big(B_i^{\mathcal{F},k}(h), S_i(h)\big)$ if and only if it is rational in $\big(B_i^{\mathcal{F},k}(h), D_i^{\mathcal{F},k-1}(h)\big)$.*

Proof. Necessity is straightforward, i.e., if s_i is rational in $\big(B_i^{\mathcal{F},k}(h), S_i(h)\big)$, then it is obviously the case that $s_i \in D_i^{\mathcal{F},k-1}(h)$ and moreover it is rational in the decision problem $\big(B_i^{\mathcal{F},k}(h), D_i^{\mathcal{F},k-1}(h)\big)$. Now, let us now prove sufficiency. Take an arbitrary $s_i \in D_i^{\mathcal{F},k-1}(h)$ and assume that it is rational in $\big(B_i^{\mathcal{F},k}(h), D_i^{\mathcal{F},k-1}(h)\big)$. Then, by definition, there exists some $\beta_i^h \in \Delta\big(B_i^{\mathcal{F},k}(h)\big)$ such that

$$U_i^h(s_i, \beta_i^h) \geq U_i^h(s_i', \beta_i^h) \tag{A.1}$$

for all $s_i' \in D_i^{\mathcal{F},k-1}(h)$. Now, assume – contrary to what we want to show – that s_i is not rational in $\big(B_i^{\mathcal{F},k}(h), S_i(h)\big)$, and take an arbitrary rational strategy s_i'' given β_i^h. Thus, it is the case that

$$U_i^h(s_i'', \beta_i^h) > U_i^h(s_i, \beta_i^h). \tag{A.2}$$

Notice that the last inequality is strict, because otherwise s_i would have been a rational strategy in $\big(B_i^{\mathcal{F},k}(h), S_i(h)\big)$. Moreover, from the previous step it follows that $s_i'' \in D_i^{\mathcal{F},k-1}(h)$. But then, this contradicts the fact that s_i is rational in $\big(B_i^{\mathcal{F},k}(h), D_i^{\mathcal{F},k-1}(h)\big)$, thus implying that s_i must necessarily be rational in $\big(B_i^{\mathcal{F},k}(h), S_i(h)\big)$. □

Now, let $T_i^{\mathcal{F},k} := \bigcap_{h \in H_i} T_i^{\mathcal{F},k}(h)$. Then, fix an arbitrary $G \in \mathcal{H} := 2^H \setminus \{\emptyset\}$, and define

$$\begin{aligned} D_i^{\mathcal{F},k}(G) &:= \{s_i \in S_i : s_i \in D_i^{\mathcal{F},k}(h) \text{ for all } h \in H_i(s_i) \cap G\} && \text{(A.3)} \\ R_i^{\mathcal{F},k}(G) &:= \{s_i \in S_i : \text{there is } t_i \in T_i^{\mathcal{F},k} \text{ such that } (s_i, t_i) \in R_i^h \text{ for all } h \in H_i(s_i) \cap G\} \\ &= \operatorname{Proj}_{S_i}\Big(R_i^G \cap \big(S_i \times T_i^{\mathcal{F},k}\big)\Big). && \text{(A.4)} \end{aligned}$$

Then, we define the set of i's strategies that survive $\mathcal{F}$-ICDP at all histories in G by

$$D_i^{\mathcal{F}}(G) := \bigcap_{k=1}^{\infty} D_i^{\mathcal{F},k}(G).$$

Likewise, we define the set of i's strategies that are rational given some type that satisfies $\mathcal{F}$-CSBR at all histories in G by

$$R_i^{\mathcal{F}}(G) := \bigcap_{k=1}^{\infty} R_i^{\mathcal{F},k}(G).$$

Construction of conditional beliefs. Fix an arbitrary $G \in \mathcal{H}$ and an arbitrary $s_i \in D_i^{\mathcal{F},1}(G)$. Then, it follows directly from Pearce (1984, Lem. 3) that for every $h \in H_i(s_i) \cap G$ there exists at least one conditional belief $\beta_{s_i,G}^h \in \Delta\big(S_{-i}(h)\big)$ such that

$$U_i^h(s_i, \beta_{s_i,G}^h) \geq U_i^h(s_i', \beta_{s_i,G}^h) \tag{A.5}$$

for all $s_i' \in S_i(h)$. Now, consider the following two cases:

- Suppose there exists some $k \in \mathbb{N}$ such that $s_i \in D_i^{\mathcal{F},k}(G) \setminus D_i^{\mathcal{F},k+1}(G)$. Then, it follows by definition that s_i is rational in $\big(B_i^{\mathcal{F},k}(h), D_i^{\mathcal{F},k-1}(h)\big)$. Hence, it follows from the optimality principle (Lemma A1) that we can choose some $\beta_{s_i,G}^h \in \Delta\big(B_i^{\mathcal{F},k}(h)\big)$ satisfying Eq. (A.5).

- Suppose that $s_i \in D_i^{\mathcal{F},k}(G)$ for all $k \in \mathbb{N}$. Then, it follows by definition that s_i is rational in $\big(B_i^{\mathcal{F},k}(h), D_i^{\mathcal{F},k-1}(h)\big)$ for every $k \in \mathbb{N}$. Thus we can choose some $\beta_{s_i,G}^h \in \Delta\big(B_i^{\mathcal{F}}(h)\big)$ satisfying Eq. (A.5).

In either of the two cases, complete the collection of conditional beliefs $\big(\beta_{s_i,G}^h\big)_{h\in H_i}$ by considering arbitrary conditional beliefs $\beta_{s_i,G}^{h'} \in \Delta\big(S_{-i}(h')\big)$ for every $h' \in H_i \setminus \big(H_i(s_i) \cap G\big)$.

Construction of types. For each player $i \in I$, define the finite set $\Theta_i := \{\theta_{s_i,G} \mid (s_i, G) \in S_i \times \mathcal{H}\}$, and let $\Theta_{-i} := \bigtimes_{j\neq i} \Theta_j$. Now, define the mapping $g_i^h : \Theta_i \to \Delta\big(S_{-i}(h) \times \Theta_{-i}\big)$ for each $h \in H_i$ as follows: For each $s_i \in D_i^1(G)$, let

$$g_i^h(\theta_{s_i,G})(s_{-i}, \theta_{-i}) := \begin{cases} \beta_{s_i,G}^h(s_{-i}) & \text{if } \theta_j = \theta_{s_j,F_i(h)} \text{ for all } j \neq i \\ 0 & \text{otherwise.} \end{cases} \tag{A.6}$$

On the other hand, if $s_i \notin D_i^{\mathcal{F},1}(G)$, let $g_i^h(\theta_{s_i,G})(s_{-i}, \theta_{-i})$ be an arbitrary probability measure over $S_{-i}(h)\times\Theta_{-i}$. Now, observe that $\big((\Theta_i)_{i\in I}, (g_i)_{i\in I}\big)$ is a finite type structure, implying that each $\theta_i \in \Theta_i$ is associated with a hierarchy of conditional beliefs.

Recall that we have assumed $\big((T_i)_{i\in I}, (\lambda_i)_{i\in I}\big)$ to be a complete type structure. Then, it follows from Friedenberg (2010) that for every hierarchy of conditional beliefs there is a type inducing this hierarchy (see Footnote 8). Thus, there is a function $\xi_i : \Theta_i \to T_i$ mapping each $\theta_{s_i,G}$ to the (unique) type $t_{s_i,G} := \xi_i(\theta_{s_i,G})$ that induces the same hierarchy in $\big((T_i)_{i\in I}, (\lambda_i)_{i\in I}\big)$. Moreover, notice that by construction it is the case that $\lambda_i^h(t_{s_i,G})(s_{-i}, t_{-i}) = g_i^h(\theta_{s_i,G})\big(s_{-i}, \xi_i^{-1}(t_{-i})\big)$. Furthermore, by construction it is the case that $(s_i, t_{s_i,G}) \in R_i^G$ whenever $s_i \in D_i^{\mathcal{F},1}(G)$.

Before we move on, for notation simplicity, let us adopt the convention that $T_i^{\mathcal{F},0}(h) := T_i$.

Lemma A2. *For every player $i \in I$, every $G \in \mathcal{H}$ and every $k > 0$, the following hold:*

(i) If $t_i \in T_i^{\mathcal{F},k-1}(h)$ then $b_i^h(t_i) \in \Delta\big(B_i^{\mathcal{F},k}(h)\big)$.

(ii) If $s_i \in D_i^{\mathcal{F},k}(G)$ then $t_{s_i,G} \in T_i^{\mathcal{F},k-1}(h)$ for all $h \in H_i(s_i) \cap G$.

(iii) $R_i^{\mathcal{F},k-1}(G) = D_i^{\mathcal{F},k}(G)$.

Proof. We prove the result by induction on k.

Initial step. First, it is rather trivial to prove the result for $k = 1$. Indeed, observe that by construction it is the case that $B_i^{\mathcal{F},1}(h) = B_i^{\mathcal{F},0}(h) = S_{-i}(h)$, and therefore $\Delta\big(B_i^{\mathcal{F},1}(h)\big) = \Delta\big(S_{-i}(h)\big)$, thus implying that $b_i^h(t_i) \in \Delta\big(B_i^{\mathcal{F},1}(h)\big)$ for all $t_i \in T_i$, which proves (i). Moreover, recall from our convention that $T_i^{\mathcal{F},0}(h) = T_i$, thus implying that $t_{s_i,G} \in T_i^{\mathcal{F},0}(h)$ for all $h \in H_i(s_i) \cap G$, irrespective of whether $s_i \in D_i^{\mathcal{F},1}$ or not, which proves (ii). Finally, notice that

$$\begin{aligned} R_i^{\mathcal{F},0}(G) &= \{s_i \in S_i : \text{there is } t_i \in T_i^{\mathcal{F},0}(h) \text{ such that } (s_i,t_i) \in R_i^h \text{ for all } h \in H_i(s_i) \cap G\} \\ &= \{s_i \in S_i : \text{there is } t_i \in T_i \text{ such that } (s_i,t_i) \in R_i^G\} \\ &= D_i^{\mathcal{F},1}(G) \end{aligned}$$

which proves (iii).

Inductive step. We assume that the result holds for an arbitrary $k > 0$. We will refer to this as our "induction assumption (IA)". Then, we are going to prove it for $k+1$.

Proof of (i): Fix some $h \in H_i$, and assume that $t_i \in T_i^{\mathcal{F},k}(h)$. Then, by definition it is the case that

$$t_i \in SB_i^h\big(R_{-i}^{F_i(h)} \cap (S_{-i} \times T_{-i}^{\mathcal{F},k-1}(F_i(h)))\big).$$

Then, we consider the following two cases:

(a) Let $R_{-i}^{F_i(h)} \cap (S_{-i} \times T_{-i}^{\mathcal{F},k-1}(F_i(h))) \neq \emptyset$.

By the definition of strong belief (at h) it is the case that $\lambda_i^h(t_i)\big(R_{-i}^{F_i(h)} \cap (S_{-i} \times T_{-i}^{\mathcal{F},k-1}(h))\big) = 1$. Now, recall by Eq. (A.4) that

$$R_{-i}^{\mathcal{F},k-1}(F_i(h)) = \mathrm{Proj}_{S_{-i}}\big(R_{-i}^{F_i(h)} \cap (S_{-i} \times T_{-i}^{\mathcal{F},k-1}(F_i(h)))\big),$$

and therefore it follows that $b_i^h(t_i)\big(R_{-i}^{\mathcal{F},k-1}(F_i(h))\big) = 1$. Now observe that

$$\begin{aligned} R_{-i}^{\mathcal{F},k-1}\big(F_i(h)\big) &= \bigtimes_{j \neq i} \left\{ s_j \in S_j : s_j \in R_j^{\mathcal{F},k-1}\big(F_i(h)\big) \right\} \\ &= \bigtimes_{j \neq i} \left\{ s_j \in S_j : s_j \in D_j^{\mathcal{F},k}\big(F_i(h)\big) \right\} && \big(\text{by the IA}\big) \\ &= \bigtimes_{j \neq i} \left\{ s_j \in S_j : s_j \in D_j^{\mathcal{F},k}(h') \text{ for all } h' \in H_j \cap F_i(h) \right\}. && \text{(A.7)} \end{aligned}$$

Thus, it is the case that

$$\begin{aligned} C_i^{\mathcal{F},k}(h) &= \bigtimes_{j \neq i} \left\{ s_j \in S_j(h) : s_j \in D_j^{\mathcal{F},k}(h') \text{ for all } h' \in H_j \cap F_i(h) \right\} \\ &= S_{-i}(h) \cap R_{-i}^{\mathcal{F},k-1}\big(F_i(h)\big). && \text{(A.8)} \end{aligned}$$

Now, there are two possibilities. According to the first possibility we have $C_i^{\mathcal{F},k}(h) \neq \emptyset$, in which case we obtain

$$\begin{aligned} B_i^{\mathcal{F},k+1}(h) &= C_i^{\mathcal{F},k}(h) \\ &= S_{-i}(h) \cap R_{-i}^{\mathcal{F},k-1}\big(F_i(h)\big). \end{aligned}$$

Then, by combining $b_i^h(t_i)\big(R_{-i}^{\mathcal{F},k-1}(F_i(h))\big) = 1$ with $b_i^h(t_i)\big(S_{-i}(h)\big) = 1$, it is straightforward to obtain $b_i^h(t_i)\big(B_i^{\mathcal{F},k+1}(h)\big) = 1$. According to the second possibility we have $C_i^{\mathcal{F},k}(h) = \emptyset$, in which case we obtain $B_i^{\mathcal{F},k+1}(h) = B_i^{\mathcal{F},k}(h)$. But then, since $t_i \in T_i^{\mathcal{F},k}(h) \subseteq T_i^{\mathcal{F},k-1}(h)$, it follows from the IA that $b_i^h(t_i)\big(B_i^{\mathcal{F},k+1}(h)\big) = b_i^h(t_i)\big(B_i^{\mathcal{F},k}(h)\big) = 1$, which completes this part of the proof.

(b) Let $R_{-i}^{F_i(h)} \cap \big(S_{-i} \times T_{-i}^{\mathcal{F},k-1}(F_i(h))\big) = \emptyset$.

Then, it follows by definition that

$$\begin{aligned} R_{-i}^{\mathcal{F},k-1}\big(F_i(h)\big) \cap S_{-i}(h) &\subseteq R_{-i}^{\mathcal{F},k-1}\big(F_i(h)\big) \\ &= \mathrm{Proj}_{S_{-i}}\Big(R_{-i}^{F_i(h)} \cap (S_{-i} \times T_{-i}^{k-1}(F_i(h)))\Big) \\ &= \emptyset \end{aligned} \tag{A.9}$$

Now, using the same reasoning as in Eq. (A.7), combined with Eq. (A.9), we obtain

$$\begin{aligned} R_{-i}^{\mathcal{F},k-1}(F_i(h)) \cap S_{-i}(h) &= \bigtimes_{j \neq i} \big\{\, s_j \in S_j(h) : s_j \in D_j^{\mathcal{F},k}(h') \text{ for all } h' \in H_j \cap F_i(h) \,\big\} \\ &= \emptyset. \end{aligned}$$

Moreover, using the same argument as in Eq. (A.8), we obtain

$$C_i^{\mathcal{F},k}(h) = S_{-i}(h) \cap R_{-i}^{\mathcal{F},k-1}\big(F_i(h)\big).$$

Thus, combining the previous two equations, we conclude that $C_i^{\mathcal{F},k}(h) = \emptyset$. Hence, $B_i^{\mathcal{F},k+1}(h) = B_i^{\mathcal{F},k}(h)$. Finally, since $t_i \in T_i^{\mathcal{F},k}(h) \subseteq T_i^{\mathcal{F},k-1}(h)$, it follows from the IA that $b_i^h(t_i)\big(B_i^{\mathcal{F},k+1}(h)\big) = b_i^h(t_i)\big(B_i^{\mathcal{F},k}(h)\big) = 1$, which completes the proof of part (i).

Proof of (ii): Take an $s_i \in D_i^{\mathcal{F},k+1}(G)$, and consider some $h \in H_i(s_i) \cap G$. Since $D_i^{\mathcal{F},k+1}(G) \subseteq D_i^{\mathcal{F},k}(G)$, it follows by the IA that $t_{s_i,G} \in T_i^{\mathcal{F},k-1}(h)$. Hence, it suffices to prove that

$$t_{s_i,G} \in SB_i^h\Big(R_{-i}^{F_i(h)} \cap \big(S_{-i} \times T_{-i}^{\mathcal{F},k-1}(h)\big)\Big). \tag{A.10}$$

The latter amounts to proving that

$$\Big(R_{-i}^{F_i(h)} \cap \big(S_{-i}(h) \times T_{-i}^{\mathcal{F},k-1}(h)\big) \neq \emptyset\Big) \;\Rightarrow\; \Big(\lambda_i^h(t_{s_i,G})\big(R_{-i}^{F_i(h)} \cap \big(S_{-i} \times T_{-i}^{\mathcal{F},k-1}(h)\big)\big) = 1\Big). \tag{A.11}$$

First, notice that $t_{s_i,G} \in SB_i^h\big(R_{-i}^{F_i(h)} \cap \big(S_{-i} \times T_{-i}^{\mathcal{F},k-1}(h)\big)\big)$ is trivially satisfied whenever $R_{-i}^{F_i(h)} \cap \big(S_{-i}(h) \times T_{-i}^{\mathcal{F},k-1}(h)\big) = \emptyset$. Hence, we will focus on the case where $R_{-i}^{F_i(h)} \cap \big(S_{-i}(h) \times T_{-i}^{\mathcal{F},k-1}(h)\big) \neq \emptyset$. For every $j \neq i$, there exists some $(s_j^*, t_j^*) \in S_j(h) \times T_j$ such that (1) $(s_j^*, t_j^*) \in R_j^{h'}$ for all $h' \in H_j(s_j^*) \cap F_i(h)$, and (2) $t_j^* \in T_j^{\mathcal{F},k-1}(h')$ for all $h' \in H_j \cap F_i(h)$.

Now, we are going to prove that $s_j^* \in D_j^{\mathcal{F},k}(h')$ for every $h' \in H_j(s_j^*) \cap F_i(h)$. To do so, take an arbitrary $t_j^{k-1} \in T_j^{\mathcal{F},k-1}$, and define the type t_j^{**} by

$$\lambda_j^{h'}(t_j^{**}) := \begin{cases} \lambda_j^{h'}(t_j^*) & \text{for each } h' \in H_j(s_j^*) \cap F_i(h), \\ \lambda_j^{h'}(t_j^{k-1}) & \text{for each } h' \in H_j \setminus \big(H_j(s_j^*) \cap F_i(h)\big). \end{cases}$$

Notice that since $\big((T_i)_{i\in I}, (\lambda_i)_{i\in I}\big)$ is a complete type structure, such a type exists. Observe that by construction it is the case that $(s_j^*, t_j^{**}) \in R_j^{F_i(h)}$, and moreover $t_j^{**} \in T_j^{\mathcal{F},k-1}$. Therefore, we obtain

$$\begin{aligned} s_j^* &\in R_j^{\mathcal{F},k-1}\big(F_i(h)\big) \cap S_j(h) \\ &= D_j^{\mathcal{F},k}\big(F_i(h)\big) \cap S_j(h) && \big(\text{by the IA}\big) \\ &= \big\{\ s_j \in S_j(h) : s_j \in D_j^{\mathcal{F},k}(h') \text{ for all } h' \in H_j(s_j) \cap F_i(h)\ \big\} \\ &\neq \emptyset. \end{aligned}$$

The latter implies directly by definition that $C_i^{\mathcal{F},k}(h) \neq \emptyset$. Hence, it is – also by definition – the case that

$$B_i^{\mathcal{F},k+1}(h) = C_i^{\mathcal{F},k}(h). \tag{A.12}$$

Now, notice that by construction $\lambda_i^h(t_{s_i,G})$ put positive probability only to strategy-type pairs (s_j, t_j) such that $t_j = t_{s_j,F_i(h)}$. Moreover, since $s_i \in D_i^{\mathcal{F},k+1}(G)$ and $t_{s_i,G} \in T_i^{\mathcal{F},k-1}(h)$ it follows from Part (i) of our result – which we have already proven above – that $b_i^h(t_{s_i,G}) \in \Delta\big(B_i^{\mathcal{F},k}(h)\big)$. Therefore, it follows from Eq. (A.12) that $\text{marg}_{S_j \times T_j} \lambda_i^h(t_{s_i,G})$ puts positive probability only to strategy-type pairs $(s_j, t_j) \in S_j(h) \times T_j$ such that $t_j = t_{s_j,F_i(h)}$ and $s_j \in D_j^{\mathcal{F},k}(h')$ for all $h' \in H_j(s_j) \cap F_i(h)$. Hence, from the IA it follows that $\text{marg}_{S_j \times T_j} \lambda_i^h(t_{s_i,G})$ assigns probability 1 to

$$R_j^{F_i(h)} \cap \big\{\ (s_j, t_j) \in S_j \times T_j : t_j \in T_j^{\mathcal{F},k-1}(h') \text{ for all } h' \in H_j \cap F_i(h)\ \big\}$$

for every $j \neq i$. Therefore, by definition, $t_{s_i,G} \in T_i^{\mathcal{F},k}(h)$, which completes the proof of part (ii).

Proof of (iii): First, we prove that $R_i^{\mathcal{F},k-1}(G) \subseteq D_i^{\mathcal{F},k}(G)$: Take an arbitrary $s_i \in R_i^{\mathcal{F},k-1}(G)$. By definition there exists a type in $t_i \in T_i^{\mathcal{F},k-1}$ such that $(s_i, t_i) \in R_i^G$. Now, by part (i) of the result – that we have already proven above – it follows that $b_i^h(t_i)\big(B_i^{\mathcal{F},k}(h)\big) = 1$ for all $h \in H_i(s_i) \cap G$, implying that at all histories $h \in H_i(s_i) \cap G$, the strategy s_i is rational in the decision problem $\big(B_i^{\mathcal{F},k}(h), D_i^{\mathcal{F},k-1}(h)\big)$. Thus, we conclude that $s_i \in D_i^{\mathcal{F},k}(h)$ for all $h \in H_i(s_i) \cap G$. The latter directly implies that $s_i \in D_i^{\mathcal{F},k}(G)$ which completes this part of the proof.

Second, we prove that $D_i^{\mathcal{F},k}(G) \subseteq R_i^{\mathcal{F},k-1}(G)$: Take an arbitrary $s_i \in D_i^{\mathcal{F},k}(G)$. Then, by part (ii) that we have already proven above, it follows that $t_{s_i,G} \in T_i^{\mathcal{F},k-1}(h)$ for all $h \in G \cap H_i(s_i)$. Now, fix an arbitrary type $t_i^{k-1} \in T_i^{\mathcal{F},k-1}$, and define the type $t^*_{s_i,G} \in T_i$ by

$$\lambda_i^h(t^*_{s_i,G}) := \begin{cases} \lambda_i^h(t_{s_i,G}) & \text{for each } h \in H_i(s_i) \cap G, \\ \lambda_i^h(t_i^{k-1}) & \text{for each } h \in H_i \setminus \big(H_i(s_i) \cap G\big). \end{cases}$$

Notice that since $\big((T_i)_{i\in I}, (\lambda_i)_{i\in I}\big)$ is a complete type structure, such a type exists. Then, by construction it is the case that $t^*_{s_i,G} \in T_i^{\mathcal{F},k-1}$, and therefore it follows that $(s_i, t^*_{s_i,G}) \in R_i^h$ for all $h \in G \cap H_i(s_i)$. Hence, we conclude that $s_i \in R_i^{\mathcal{F},k-1}(G)$, which completes the proof. □

Proof of Theorem 1. Take an arbitrary $i \in I$ and some $h \in H_i$.

Proof of (i): It follows directly from Lemma A2.i.

Proof of (ii): Fix an arbitrary $\beta_i^h \in \Delta\big(B_i^{\mathcal{F},k}(h)\big)$, and let $s_i^* \in D_i^{\mathcal{F},k}(h)$ be such that

$$U_i^h(s_i^*, \beta_i^h) \geq U_i^h(s_i, \beta_i^h) \tag{A.13}$$

for all $s_i \in D_i^{\mathcal{F},k-1}(h)$. In fact, notice that Eq. (A.13) holds, not only for every $s_i \in D_i^{\mathcal{F},k-1}(h)$, but for every $s_i \in S_i(h)$ (see Lemma A1). Now, we define $\beta^h_{s_i^*,\{h\}} := \beta_i^h$, and construct the type $t^h_{s_i^*,\{h\}}$ like we did above. Then, by Lemma A2.ii, it is the case that $t_{s_i^*,\{h\}} \in T_i^{\mathcal{F},k-1}(h)$, which – together with the fact that $\beta^h_{s_i^*,\{h\}} := b_i^h(t_{s_i^*,\{h\}})$ – completes the proof. □

Proof of Theorem 2. Observe that by construction

$$\begin{aligned} R_i^{\mathcal{F}}(H) &= \mathrm{Proj}_{S_i}\big(R_i \cap (S_i \times T_i^{\mathcal{F}})\big) \\ D_i^{\mathcal{F}}(H) &= \{s_i \in S_i : s_i \in D_i^{\mathcal{F}}(h) \text{ for all } h \in H_i(s_i)\}, \end{aligned}$$

and recall by Lemma A2.iii that $R_i^{\mathcal{F}}(H) = D_i^{\mathcal{F}}(H)$, which completes the proof. □

B. Proofs of Section 4

Proof of Proposition 1. We proceed by induction on k. First, note that $SB_i^1 = \bigcap_{h\in H_i} T_i^{\mathcal{F},1}(h)$. Then, assume that for every $i \in I$ it is the case that $SB_i^{k-1} = \bigcap_{h\in H_i} T_i^{\mathcal{F},k-1}(h)$. Now, observe that for every $i \in I$ and $h \in H_i$, it is the case that

$$\begin{aligned} T_{-i}^{\mathcal{F},k-1}\big(F_i(h)\big) &= \bigtimes_{j\neq i} \{\, t_j \in T_j : t_j \in T_j^{\mathcal{F},k-1}(h') \text{ for all } h' \in H_j\} \\ &= \bigtimes_{j\neq i} \Big(\bigcap_{h'\in H_j} T_j^{\mathcal{F},k-1}(h')\Big) \\ &= \bigtimes_{j\neq i} SB_j^{k-1} \\ &= SB_{-i}^{k-1}. \end{aligned}$$

Hence, it is the case that

$$\begin{aligned} SB_i^k &= SB_i^{k-1} \cap SB_i\big(R_{-i} \cap (S_{-i} \times SB_{-i}^{k-1})\big) \\ &= \Big(\bigcap_{h \in H_i} T_i^{\mathcal{F},k-1}(h)\Big) \cap \Big(\bigcap_{h \in H_i} SB_i^h\Big(R_{-i}^{F_i(h)} \cap \big(S_{-i} \times T_{-i}^{\mathcal{F},k-1}\big(F_i(h)\big)\big)\Big)\Big) \\ &= \bigcap_{h \in H_i} \Big(T_i^{\mathcal{F},k-1}(h) \cap SB_i^h\Big(R_{-i}^{F_i(h)} \cap \big(S_{-i} \times T_{-i}^{\mathcal{F},k-1}\big(F_i(h)\big)\big)\Big)\Big) \\ &= \bigcap_{h \in H_i} T_i^{\mathcal{F},k}(h) \end{aligned}$$

which completes the proof. □

In order to prove Proposition 2, we first recall the formal definition of the backward dominance procedure (BDP), originally introduced by Perea (2014).

Backward dominance procedure. For an arbitrary $i \in I$ and an arbitrary $h \in H$, consider the following sequence of subsets of $S_i(h)$:

$$\begin{aligned} Q_i^1(h) &:= S_i(h) \\ Q_i^2(h) &:= \{s_i \in Q_i^1(h) : s_i \text{ is rational in } \big(Q_{-i}^1(h'), Q_i^1(h')\big) \text{ at all } h' \in H_i(s_i) \cap \mathrm{Fut}(h)\} \\ &\ \vdots \\ Q_i^k(h) &:= \{s_i \in Q_i^{k-1}(h) : s_i \text{ is rational in } \big(Q_{-i}^{k-1}(h'), Q_i^{k-1}(h')\big) \text{ at all } h' \in H_i(s_i) \cap \mathrm{Fut}(h)\} \\ &\ \vdots \end{aligned}$$

for each $k > 0$, where $Q_{-i}^k(h) = \times_{j \neq i} Q_j^k(h)$. We say that a strategy s_i survives k steps of the procedure at $h \in H_i$ whenever $s_i \in Q_i^k(h)$. The idea is that a strategy survives k steps of the procedure at some $h \in H_i$ whenever it is not strictly dominated in the corresponding normal form game – that has survived so far – at every history following h where i is active. Then, we define

$$Q_i(h) := \bigcap_{k=1}^{\infty} Q_i^k(h), \tag{B.1}$$

and we say that a strategy survives the procedure whenever it is the case that $s_i \in Q_i(h)$ for all $h \in H_i(s_i)$.

Now, let us prove an intermediate lemma that we will use in the proof of Proposition 2.

Lemma B1. *Let $\mathcal{F}$ be such that $F_i(h) = \mathrm{Fut}_{-i}(h)$ for every $i \in I$ and every $h \in H_i$. Then, for every $i \in I$, every $h \in H_i$ and every $k > 1$ the following hold:*

(i) $Q_{-i}^k(h) = B_i^{\mathcal{F},k}(h)$.

(ii) $Q_i^{k+1}(h) = \{s_i \in S_i(h) : s_i \in D_i^{\mathcal{F},k}(h') \text{ for all } h' \in \mathrm{Fut}(h) \cap H_i(s_i)\}$.

Proof. We proceed to prove the result by induction on k. The result trivially holds for $k = 1$. We assume it holds for $k-1$ and we will prove it for k. We begin with part (i). Fix an arbitrary $i \in I$ and an arbitrary $h \in H_i$, and observe that

$$\begin{aligned} B_i^{\mathcal{F},k}(h) &= C_i^{\mathcal{F},k-1}(h) \\ &= \bigtimes_{j \neq i} \{s_j \in S_j(h) : s_j \in D_j^{\mathcal{F},k-1}(h') \text{ for all } h' \in H_j(s_j) \cap \mathrm{Fut}(h)\} \\ &= \bigtimes_{j \neq i} Q_j^k(h) && \big(\text{by the IA}\big) \\ &= Q_{-i}^k(h), \end{aligned}$$

which completes the inductive step of the proof for part (i).

Now, we move the inductive step for part (ii). Again, fix an arbitrary $i \in I$ and an arbitrary $h \in H_i$, and take an arbitrary $s_i \in Q_i^{k+1}(h)$. Then, by definition, s_i is rational in $\big(Q_{-i}^k(h'), Q_i^k(h')\big)$ for every $h' \in \mathrm{Fut}(h) \cap H_i(s_i)$, and by part (i) of the present result, s_i is rational in $\big(B_i^{\mathcal{F},k}(h'), Q_i^k(h')\big)$ for every $h' \in \mathrm{Fut}(h) \cap H_i(s_i)$. Now, notice that for every $s_i' \in S_i(h')$,

$$\begin{aligned} \text{a strategy } s_i' \text{ is rational in } \big(B_i^{\mathcal{F},k}(h'), Q_i^k(h')\big) \quad &\Leftrightarrow \quad \text{a strategy } s_i' \text{ is rational in } \big(B_i^{\mathcal{F},k}(h'), S_i(h')\big) \\ &\Leftrightarrow \quad \text{a strategy } s_i' \text{ is rational in } \big(B_i^{\mathcal{F},k}(h'), D_i^{\mathcal{F},k-1}(h')\big). \end{aligned}$$

The first equivalence follows from Perea (2012, Lem. 8.14.6), while the second one follows from Lemma A1. Hence, s_i is rational in $\big(B_i^{\mathcal{F},k}(h'), D_i^{\mathcal{F},k-1}(h')\big)$ for every $h' \in \mathrm{Fut}(h) \cap H_i(s_i)$, thus implying that $s_i \in D_i^{\mathcal{F},k}(h')$ for every $h' \in \mathrm{Fut}(h) \cap H_i(s_i)$. Therefore,

$$Q_i^{k+1}(h) \subseteq \{s_i \in S_i(h) : s_i \in D_i^{\mathcal{F},k}(h') \text{ for all } h' \in \mathrm{Fut}(h) \cap H_i(s_i)\}. \tag{B.2}$$

Now, in order to prove the inverse weak inequality, take some $s_i \in D_i^{\mathcal{F},k}(h')$ for every $h' \in \mathrm{Fut}(h) \cap H_i(s_i)$. This implies that s_i is rational in $\big(Q_{-i}^k(h'), D_i^{\mathcal{F},k-1}(h')\big)$ for every $h' \in \mathrm{Fut}(h) \cap H_i(s_i)$, and by the previous sequence of equivalences, s_i is rational in $\big(Q_{-i}^k(h'), Q_i^k(h')\big)$ for every $h' \in \mathrm{Fut}(h) \cap H_i(s_i)$. Then, by definition, $s_i \in Q_i^{k+1}(h)$, thus proving that

$$Q_i^{k+1}(h) \supseteq \{s_i \in S_i(h) : s_i \in D_i^{\mathcal{F},k}(h') \text{ for all } h' \in \mathrm{Fut}(h) \cap H_i(s_i)\}. \tag{B.3}$$

Then, inequalities (B.2) and (B.3) complete this part of the proof. □

Proof of Proposition 2. It follows from Perea (2014, Thm. 5.4) that a strategy can be rationally played under CBFR (in a complete type structure) if and only if it survives the BDP, i.e., formally, $s_i \in Q_i(h)$ for all $h \in H_i(s_i)$ if and only if $s_i \in \mathrm{Proj}_{S_i}\big(R_i \cap (S_i \times CFB_i)\big)$. Moreover, from our Theorem 2, a strategy s_i can be rationally played under $\mathcal{F}$-CSBR (in a complete type structure) if and only if it survives the $\mathcal{F}$-ICDP, i.e., formally, $s_i \in D_i^{\mathcal{F}}(h)$ for all $h \in H_i(s_i)$ if and only if

$s_i \in \mathrm{Proj}_{S_i}\big(R_i \cap (S_i \times T_i^{\mathcal{F}})\big)$. Thus, it suffice to prove that a strategy survives BDP if and only if it survives $\mathcal{F}$-ICDP.

First, consider an arbitrary strategy s_i surviving the BDP. Then, it must be the case that $s_i \in Q_i^k(h)$ for every $k > 0$ and every $h \in H_i(s_i)$. Thus, by Lemma B1, the latter is true if and only if $s_i \in \{s_i' \in S_i(h) : s_i' \in D_i^{\mathcal{F},k}(h') \text{ for all } h' \in \mathrm{Fut}(h) \cap H_i(s_i)\}$ for all $k > 0$ and for all $h \in H_i(s_i)$. Obviously, the latter is equivalent to $s_i \in D_i^{\mathcal{F},k}(h)$ for every $k > 0$ and every $h \in H_i(s_i)$, which by definition means that s_i survives the $\mathcal{F}$-ICDP, thus completing the proof. □

References

Baltag, A., Smets, S. & Zvesper, J.A. (2009). Keep hoping for rationality: a solution to the backward induction paradox. *Synthese* 169, 705–737.

Battigalli, P. (1997). On rationalizability in extensive games. *Journal of Economic Theory* 74, 40-61.

Battigalli, P. & Friedenberg, A. (2012). Forward induction reasoning revisited. *Theoretical Economics* 57–98.

Battigalli, P. & Siniscalchi, M. (1999). Hierarchies of conditional beliefs and interactive epistemology in dynamic games. *Journal of Economic Theory* 88, 188–230.

——— (2002). Strong belief and forward induction reasoning. *Journal of Economic Theory* 106, 356–391.

Chen, J. & Micali, S. (2013). The order independence of iterated dominance in extensive games. *Theoretical Economics* 8, 125163.

Cho, I.K. (1987). A refinement of sequential equilibrium. *Econometrica* 55, 1367–1389.

Cho, I.K. & Kreps, D.M. (1987). Signaling games and stable equilibria. *Quarterly Journal of Economics* 102, 179–221.

Friedenberg, A. (2010). When do type structures contain all hierarchies of beliefs? *Games and Economic Behavior* 68, 108–129.

Govindan, S. & Wilson, R. (2009). On forward induction. *Econometrica* 77, 1–28.

Harsanyi, J. (1967-68). Games with incomplete information played by Bayesian players, I-III. *Management Science* 14, 159–182, 320–334, 486–502.

Heifetz, A. & Perea, A. (2014). On the Outcome Equivalence of Backward Induction and Extensive Form Rationalizability. *International Journal of Game Theory* (forthcoming).

Hillas, J. (1994). Sequential equilibria and stable sets of beliefs. *Journal of Economic Theory* 64, 78–102.

Kreps, D.M. & Wilson, R. (1982). Sequential equilibria. *Econometrica* 50, 863–894.

McLennan, A. (1985). Justifiable beliefs in sequential equilibrium. *Econometrica* 53, 889–904.

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

Penta, A. (2009). Robust dynamic mechanism design. *Working paper*.

Perea, A. (2012). *Epistemic game theory: Reasoning and choice.* Cambridge University Press.

——— (2014). Belief in the opponents' future rationality. *Games and Economic Behavior* 83, 231–254.

Rênyi, A. (1955). On a new axiomatic theory of probability. *Acta Mathetica Academiae Scientiarum Hungaricae* 6, 285–335.

Rubinstein, A. (1991). Comments on the Interpretation of Game Theory. *Econometrica* 59, 909–924.

Selten, R. (1965). Spieltheoretische Behandlung eines Oligopolmodells mit Nachfragezeit. *Zeitschrift für die Gesammte Staatswissenschaft* 121, 301–324, 667–689.

Shimoji, M. & Watson, J. (1998). Conditional dominance, rationalizability, and game forms. *Journal of Economic Theory* 83, 161-195.

Stalnaker, R. (1998). Belief revision in games: Forward and backward induction. *Mathematical Social Sciences* 36, 31-56.