Local reasoning in dynamic games*

Andrés Perea†

Department of Quantitative Economics & EpiCenter, Maastricht University

Elias Tsakas‡

Department of Economics & EpiCenter, Maastricht University

December 8, 2015

Abstract

In this paper we introduce a novel framework for modeling bounded reasoning in dynamic games, based on the idea that at each history of the game each player pays attention to some – not necessarily all – histories. We refer to this phenomenon as local reasoning, and we show that several extensively-studied types of bounded rationality can be studied within this framework, such as for instance limited memory or limited foresight. Then, we proceed to study a standard form of reasoning within our framework, according to which each player tries to rationalize her opponents' past actions *at the histories that she reasons about.* As a result we obtain a generalized solution concept, which we call local common strong belief in rationality. We characterize the strategy profiles that can be rationally played under our concept by means of a simple iterative elimination procedure. Finally, we show that standard existing solution concepts – such as extensive-form rationalizability or the backward dominance procedure – are special cases of rationality and local common strong belief in rationality.

Keywords: Local reasoning, limited attention, epistemic game theory, strong belief in rationality, forward induction, backward induction, limited memory, limited foresight.

*We are indebted to Giacomo Bonanno and 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

Following the seminal work of Sims (2003), limited attention has recently attracted a lot of interest among economists. The underlying idea is that economic agents often have access to more information than they can actually process and therefore they choose to disregard certain aspects of it. This stream of research has mostly focused on understanding how the agents allocate their (limited) cognitive resources in an optimal way.

In this paper, we study limited attention from a different perspective. Instead of trying to understand *how the agents decide where to focus their attention*,[1] we study how limited attention affects their reasoning, and consequently their behavior, in dynamic settings. In particular, we ask the question of *how players choose their strategies in a dynamic game if they focus their attention on only some, but not necessarily all, histories in the game.* In fact, we have in mind dynamic games with players who disregard certain parts of the game when they form their beliefs about their opponents' behavior. Throughout the paper we refer to this phenomenon as *local reasoning*. Answering this question has important implications for economic theory in general, in fields like (epistemic) game theory and bounded rationality for instance, as well as for several specific topics, such as signaling games, long cheap talk, dynamic voting models, and of course the recently surging (in)attention literature.

Consider for instance the example of an electoral campaign where each candidate has a previous record of actions. While past actions provide valuable information to the voters regarding some candidate's future policies – if elected – it is often the case that the voters disregard some of these actions, either because they occurred a long time ago, or because they did not affect them personally, or even because they simply do not deem them of major importance.[2] In this case, the voters' beliefs, and consequently their voting behavior may be significantly affected, and moreover the candidate could in principle take this fact into account when designing her electoral campaign. In our framework this can be modelled by considering voters who do not reason about the histories at which these particular actions were undertaken by the candidate.

Perhaps of more interest to the economic theorist would be the following example that illustrates

[1] Understanding how attention is distributed optimally has been recently studied within decision-theoretic settings (e.g., De Oliveira et al., 2014; Ellis, 2015).

[2] There is extensive research within psychology indicating that the memories of the voters can even predict electoral outcomes. For an overview of this literature we refer to the textbook by Lau and Redlawsk (2006). Similar questions have also been studied by anthropologists (Cole, 2001). In all this work the notion of limited memory is interpreted as people disregarding past information that they do have access to, rather than having forgotten the information in the sense of being unaware of it, like for instance in much of the work in economics. As we are going to see in the paper, this is a crucial distinction.

how our framework could be relevant for the study of signaling games.[3] Take for instance a sender-receiver game, where Ann (the sender) sends a sequence of signals – instead of a single signal as often assumed in the literature – to Bob (the receiver), who in turn chooses an action. Then, obviously, if Bob disregards some of the earlier histories, he essentially disregards some of Ann's signals, and the aggregate information contained in the signals that he takes into account may differ from the one he would have obtained if he had been been paying attention to all of the signals. Of course, this could have behavioral implications for both the sender and the receiver.

Our formal framework is built on the general idea of each player i being exogenously endowed with a function mapping each history h to a collection of histories $F_i(h)$ that the player reasons about, upon i finding herself at the history h. For instance, in the previous example, h is the history at which the election takes place, and voter i does not reason at h about the past history $h' \notin F_i(h)$ at which the candidate j (e.g., say the former president) decided to raise taxes. Thus, i will not take j's behavior into account when forming beliefs on how j would behave (after h) in case she was elected. We refer to the function F_i as i's focus function. Then, we further enrich our framework by allowing for incomplete information about the focus functions that the players may have. The latter provides enough flexibility for our model to be able to capture certain forms of bounded reasoning that have over the past decades attracted interest among economists, such as for instance limited memory or limited foresight. In Section 2.3 it becomes clear why – technically speaking – these forms of bounded reasoning cannot be captured unless we allow for incomplete information about the players' focus function (see Ex. 2).

Thus, the first general contribution of the paper is that it provides a general novel framework for modelling limited attention in dynamic games. Before moving on, we should stress two important points. First, we do not model how players choose which histories to reason about, but rather we assume that this is exogenously given and has already been decided at some earlier stage, not included in our model. This distinguishes our work from previous literature on inattention. Second, even though the players do not reason about all histories in the game, they can still "see" the entire game. This distinguishes our work from the literature on unawareness. Both these distinctions are extensively discussed throughout the paper.

Then, using our framework, we go on to identify a number of special cases where it can be directly applied. Each special case corresponds to a specification of each player's focus function, as well as her beliefs and higher order beliefs about everybody's focus function.[4] Thus, our second general

[3] Note that signaling games assume incomplete information about the utility functions, a case we do not explicitly explore in this paper. However, as we briefly discuss in Section 2.2, our framework can be directly extended to allow for incomplete information, with all our results remaining valid.

[4] Throughout the paper, we refer to a specification of a player's focus as well as her beliefs and higher order beliefs

contribution is that we manage to embed seemingly unrelated (existing) forms of bounded reasoning – which so far have been studied independently in the literature – under the common umbrella of local reasoning, thus being able to also compare the respective behavioral predictions. Furthermore, our framework allows us to study new interesting forms of bounded reasoning in dynamic environments, such as for instance focusing exclusively on focal histories which can provide additional insights to our understanding of behavior in the existence of high (resp. low) stakes decisions (see Section 2.3).

Up to this point, we have motivated our work by stressing the novelty of our framework without having referred to the exact reasoning process that we have in mind. In other words, we have exogenously specified which histories the players reason about, but we have not described yet how the players reason about these histories. To do this, let us first recall that the product of a player's game-theoretic reasoning is typically a configuration of this players' belief hierarchy.[5] Incorporating belief hierarchies in our models has only become possible with the recent surge of the epistemic approach to game theory. Within the framework of epistemic game theory we can precisely define what it means to "believe in the opponents' rationality". Then, once we have expressed such notions, we can specify a reasoning process based on some form of common belief in rationality.

Our reasoning process is based on Battigalli and Siniscalchi's (2002) epistemic notion of strong belief in rationality. According to their original definition, a player (say Bob) strongly believes in the opponent's (say 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 the future. This already suggests that in Battigalli and Siniscalchi's (2002) framework, Bob is implicitly assumed to reason about Ann's behavior at all histories, and in this sense their model is one with global reasoning, as opposed to the local reasoning that we study here. Thus, we generalize their notion of strong belief in rationality, by introducing our concept of *local strong belief in rationality* given the players' focus. The underlying idea is that while being at h, *Bob tries to rationalize Ann's behavior, not at all histories leading to h, but rather at all histories leading to h that Bob reasons about.* More precisely, if at h it is possible for Bob to believe that Ann chooses optimally at all histories in $F_i(h)$, then he will necessarily believe (at h) that Ann indeed chooses optimally at every history in $F_i(h)$. Then, similarly to Battigalli and Siniscalchi (2002), we iterate strong belief to obtain our solution concept of *local common strong belief in rationality*, thus implying that in our

about the opponents' focus, as the player's focus type. For a detailed description of the model, see Section 2.2.

[5] 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).

model player i disregards not only her opponents' behavior at histories outside $F_i(h)$, but also their reasoning at those histories. In other words, assuming local reasoning has implications also for higher order beliefs.

Our main result provides a characterization of the strategies that can be rationally played by each player at each history under our concept by means of a simple iterative procedure. The procedure is called *local iterated conditional dominance*, thus highlighting the similarities that it bears with Shimoji and Watson's (1998) iterated conditional dominance procedure. Our procedure simultaneously eliminates strategies and conditional beliefs for each history at each step, thus inducing not only the predictions of our concept, but also the outcome of the reasoning of each player at each history. This makes our procedure a tractable tool for making predictions, and eventually testing them, especially since this is a finite procedure.

Then, we undertake our next major task, to study the relationship of our local common strong belief in rationality with standard existing solution concepts, thus placing it in its right position in the literature. 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.[6] Well-known examples of FI include extensive-form rationalizability (Pearce, 1984; Battigalli, 1997) 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, 2015).[7]

From our 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 different special cases within our framework. This unification into a generalized solution concept would also allow 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, 2015), but still the intuitive relationship of the two remains a bit unclear. In this sense, this paper constitutes a systematic attempt to close this conceptual gap. Indeed, we formally prove that stan-

[6]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.

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

dard FI and BI solution concepts are actually special cases of our 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 of Battigalli and Siniscalchi (2002). Likewise, we formally prove that the strategies that can be rationally played under Perea's (2014) 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 present and future histories.

The paper is structured as follows: In Section 2 we present our basic framework. In Section 3 we introduce the epistemic structure. In Section 4 we define the solution concept as well as the iterative conditional dominance procedure, and we present our main characterization result. In Section 5 we discuss different special cases of local reasoning. Section 6 concludes. All the proofs are relegated to the Appendices.

2. Basic framework

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. Formally, the game structure 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 set of 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 $\Pr(h)$ 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. Finally, Z denotes the set of terminal histories, i.e., the histories where no player moves, while $\Pr(z) \subseteq H$ denotes the non-terminal histories preceding z.

Moves and strategies. The finite set of moves (also 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 $\zeta : S \to Z$, mapping each strategy profile $s \in S$ to a unique terminal history. 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 $\zeta(s)$.

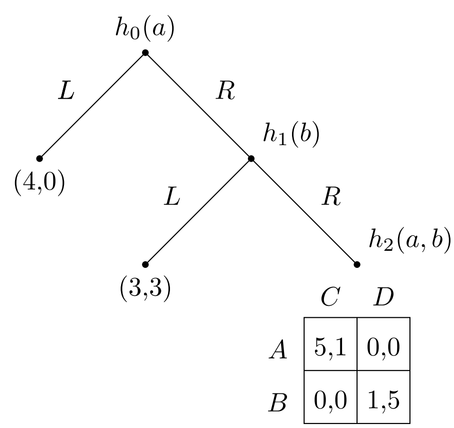


Figure 1: Generalized BoS with outside options.

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 $\zeta(s)$. Thus, we obtain the utility function $u_i : S \to \mathbb{R}$, defined as the composition $u_i := v_i \circ \zeta$, that represents i's preferences over S.[8] For instance, in Fig. 1 the strategy profile (RA, L) induces the terminal history that yields a utility of 3 to each player.

2.2. Local reasoning

In this section we introduce the notion of a player's focus (of attention) at some history. This concept is new to the literature and we will use it throughout the paper to model dynamic games with players who reason locally about some part of the game only, while disregarding the remaining histories. This tool will provide a general framework which will allow us to systematically study interesting forms of bounded rationality, such as for instance limited memory or limited foresight, just to mention a few. Moreover, some standard forms of reasoning – such as backward induction and forward induction – can be embedded as special cases within this framework.

Formally, for an arbitrary player i, the function

$$F_i : H_i \to 2^H \tag{1}$$

specifies a subset of histories that player i reasons about, upon finding herself at history $h \in H_i$.[9] Throughout the paper, we refer to F_i as i's *focus function*, with $F_i(h)$ being i's focus at h.

The general idea is that player i does not pay attention to certain histories in the game, either because she cannot, or because it is cognitively very costly, or even because she simply does not find it important to do so. In fact, the precise reason for not reasoning about certain histories may vary, depending on which these histories are. Thus, a different motivation should be provided for each special case (see Section 2.3). In either case, it is important to stress that in our model, player i is aware of the existence of all histories, including those that do not belong to $F_i(h)$.[10]

Now, let $\mathcal{F}_i$ denote the set of i's focus functions, letting $\mathcal{F} := \times_{i \in I} \mathcal{F}_i$ be the set of respective function profiles with typical element F. Following Harsanyi (1967-68) and Battigalli and Siniscalchi (1999), we model interactive uncertainty about the players' focus using an ($\mathcal{F}$-based) type structure

$$\mathfrak{F} = \big((\Theta_i)_{i \in I}, (f_i)_{i \in I}, (g_i)_{i \in I}\big),$$

where Θ_i is a finite set of focus types, also called Θ_i-types or i-agents,[11] $f_i : \Theta_i \to \mathcal{F}_i$ associates each i-agent to a focus function and $g_i : \Theta_i \times H_i \to \Delta(\Theta_{-i})$ is a function mapping each i-agent $\theta_i \in \Theta_i$ at

[8] 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.

[9] As usual, 2^H denotes the power set of H, i.e., the collection of all subsets of H.

[10] In Section 6.2 we discuss the relationship of this framework to the literature on dynamic games with unawareness.

[11] In Section 6.1 we discuss the case of infinite Θ_i's.

each history $h \in H_i$ to a conditional belief $g_i^h(\theta_i) \in \Delta(\Theta_{-i})$, with $\Theta_{-i} := \bigtimes_{j \neq i} \Theta_j$.[12]

For notation simplicity, throughout the paper we write $F_{\theta_i} := f_i(\theta_i)$. Thus, each θ_i is a full description of i's focus, as well as her beliefs and higher-order beliefs about every player's focus, and in this sense, as we are going to see later in the paper, player i's conditional beliefs about the opponents' strategies depend on the specification of θ_i.

Throughout the paper, we assume that players do not reason indirectly about a history, unless they also reason directly about this history. Formally, we impose the following assumption.

Assumption 1. For every $i \in I$, every $h \in H_i$ and every $\theta_i \in \Theta_i$, if $g_i^h(\theta_i)\big((\theta_j)_{j \neq i}\big) > 0$ then $F_{\theta_j}(h') \subseteq F_{\theta_i}(h)$ for all $h' \in F_{\theta_i}(h) \cap H_j$ and for all $j \neq i$. $\triangleleft$

Below, we present an example of a structure $\mathfrak{F}$ satisfying Assumption 1.

Example 1. Recall the game in Fig. 1, and take the structure $\mathfrak{F} = (\Theta_a, \Theta_b, f_a, f_b, g_a, g_b)$ with $\Theta_a = \{\theta_a, \theta'_a\}$ and $\Theta_b = \{\theta_b, \theta'_b\}$. Then, for each $i \in \{a, b\}$, each $\theta_i \in \Theta_i$ and each $h \in H_i$, let

$$\begin{array}{llll}
F_{\theta_a}(h_0) = H, & F_{\theta_a}(h_2) = H, & g_a^{h_0}(\theta_a) = (0.50 \otimes \theta_b\ ;\ 0.50 \otimes \theta'_b), & g_a^{h_2}(\theta_a) = (0.25 \otimes \theta_b\ ;\ 0.75 \otimes \theta'_b) \\
F_{\theta'_a}(h_0) = H, & F_{\theta'_a}(h_2) = \{h_2\}, & g_a^{h_0}(\theta'_a) = (0.35 \otimes \theta_b\ ;\ 0.65 \otimes \theta'_b), & g_a^{h_2}(\theta'_a) = (1 \otimes \theta_b) \\
F_{\theta_b}(h_1) = H, & F_{\theta_b}(h_2) = \{h_2\}, & g_b^{h_1}(\theta_b) = (0.50 \otimes \theta_a\ ;\ 0.50 \otimes \theta'_a), & g_b^{h_2}(\theta_b) = (1 \otimes \theta'_a) \\
F_{\theta'_b}(h_1) = H, & F_{\theta'_b}(h_2) = H, & g_b^{h_1}(\theta'_b) = (0.25 \otimes \theta_a\ ;\ 0.75 \otimes \theta'_a), & g_b^{h_2}(\theta'_b) = (0.50 \otimes \theta_a\ ;\ 0.50 \otimes \theta'_a)
\end{array}$$

Notice that the structure $\mathfrak{F}$ satisfies Assumption 1, e.g., θ'_a at h_2 reasons only about the present history h_2 and therefore deems θ'_b impossible, as θ'_b reasons at h_2 about histories that θ'_a herself does not, viz., $F_{\theta'_b}(h) \nsubseteq F_{\theta'_a}(h_2)$ for some $h \in F_{\theta'_a}(h_2)$. $\triangleleft$

Remark 1. In order to simplify the presentation of our results, we henceforth restrict attention to cases where the support of $g_i^h(\theta_i)$ is a singleton. Still, our analysis can be directly extended to any structure $\mathfrak{F}$.[13] Finally, throughout the paper, we are often interested in cases where the focus function of each player is transparent across all players. Formally, this is the case whenever each Θ_i is a singleton. In this case, we identify the unique $\theta_i \in \Theta_i$ with i, thus simply writing F_i for F_{θ_i}, and we say that $F \in \mathcal{F}$ is commonly known. $\triangleleft$

[12] Throughout the paper we consider only complete information games. However, our framework can be easily extended to accomodate incomplete information games, by simply introducing for each $i \in I$ a second function $\delta_i : \Theta_i \to \mathcal{U}_i$, where $\mathcal{U}_i$ is the set of i's utility functions. In this extended model we would be able to study interesting settings, like for instance signaling games.

[13] In fact, our analysis does not even depend on Assumption 1, thus implying that it could be in principle dispensed with. Still, we find it to be intuitively appealing, and as such throughout the paper we consider structures $\mathfrak{F}$ that satisfy the assumption.

2.3. Special cases

In this section we identify various special cases that can be embedded within our framework, thus illustrating its generality on the one hand, and stressing its ability to compare different forms of (perfect or bounded) reasoning in terms of their respective predictions on the other hand. We classify our special cases into two families, those corresponding to underlying reasoning structures that we encounter in the standard analysis of dynamic games in the literature, and those corresponding to different forms of bounded reasoning in dynamic games.

2.3.1. Standard forms of reasoning

Reasoning about the entire game: Forward induction. The most obvious special case is the one where it is commonly known that every player reasons at all histories about all histories. Formally, this is the case when $\Theta_i = \{\theta_i\}$ and $F_{\theta_i}(h) = F_i(h) = H$ for all $h \in H_i$ and all $i \in I$. This is in fact the underlying reasoning structure in games where the players use forward induction, according to which players look at the past before forming their beliefs about the opponents' present and future behavior. We analyze this case in detail in Section 5.1.

Reasoning about the future: Backward induction. The second, again rather obvious, special case is the one where it is common knowledge that each player reasons at all histories only about the present and the future. Formally, this is the case whenever $\Theta_i = \{\theta_i\}$ and $F_{\theta_i}(h) = F_i(h) = \mathrm{Fut}(h)$ for all $h \in H_i$ and all $i \in I$. Analogously to the previous case, this is the underlying reasoning structure in games where the players use backward induction, according to which players disregard the past before forming their beliefs about the opponents' present and future behavior. For a detailed analysis, see Section 5.2.

2.3.2. Bounded reasoning

Limited memory. Dynamic games with limited memory have been extensively studied within the game-theoretic literature.[14] Most of this earlier work has focused either on understanding the structural foundations of imperfect recall or on studying rational behavior in the existence of limited

[14] There are different strands of literature within decision theory and game theory that take into account the possibility of players exhibiting some form of limited memory. In particular, on the one hand, there is a rather large literature on how to model imperfect recall in dynamic games/decision problems and its behavioral consequences (e.g., Dow, 1991; Rubinstein, 1991, 1998; Piccione and Rubinstein, 1997; Kline, 2002; Bonanno, 2004). On the other hand, there is an also extended literature on repeated games with limited memory (e.g., Aumann and Sorin, 1989; Lehrer, 1988; Cole and Kocherlakota, 2005; Barlo et al., 2009).

memory. In either case, the common denominator in many of these papers is that players are – explicitly or implicitly – assumed to literally *not be able to remember certain past events*, viz., players forget past moves and/or past information. In this respect, players behave *as if* they cannot even "see" events that took place outside their memory horizon, and more importantly they cannot do anything about it, i.e., it is not a conscious decision they make to disregard these past events.[15] This last part is in contrast to other existing models of limited memory where the players optimally choose which pieces of information to remember, like for instance in Dow (1991).

In our framework, we consider players who can see the entire game, but still they reason only about some past histories. It is crucial to stress here that in principle our model can conceptually accommodate both previously-mentioned approaches, viz., our structure can be used to model players who cannot remember certain past events, but it can also be used to model players who consciously choose to disregard these past events, and in this sense we provide a general theory of limited memory in games. Let us now illustrate how we model limited memory in our framework.

Formally, we consider a structure $\mathfrak{F}$ such that for every $\theta_i \in \Theta_i$ and every $h \in H_i$ there exists some $h^{\leftarrow}_{\theta_i} \in \Pr(h)$ with

$$F_{\theta_i}(h) = \mathrm{Fut}(h^{\leftarrow}_{\theta_i}).$$

In this case, we refer to $h^{\leftarrow}_{\theta_i}$ as θ_i's memory horizon at h, while the number of histories separating h and $h^{\leftarrow}_{\theta_i}$ is called the memory length of θ_i at h.

Example 2. Take for instance, the game in Fig. 1, and assume that both players have a memory length of 1 throughout the entire game, i.e., at each history their memory horizon is the immediate predecessor (as long as it exists of course). Formally this means that their respective realized Θ_i-types, θ_a and θ_b, are such that $F_{\theta_a}(h_0) = H$ and $F_{\theta_a}(h_2) = \{h_1, h_2\}$ for Ann, and $F_{\theta_b}(h_1) = H$ and $F_{\theta_b}(h_2) = \{h_1, h_2\}$ for Bob. But then, Assumption 1 postulates that θ_a cannot put positive probability at h_2 to Bob's actual Θ_b-type θ_b, as $F_{\theta_b}(h_1) \nsubseteq F_{\theta_a}(h_2)$, and therefore we must introduce additional Θ_b-types that θ_a can deem possible at h_2 if we want to capture the essence of limited memory. This highlights the need for incomplete information about the players' focus. ◁

From the previous example, it becomes apparent that limited memory will eventually influence Ann's higher order beliefs – and consequently her behavior – as she does not reason at h_2 about Bob's actual Θ_b-type, but rather about some other $\theta'_b \in \Theta_b \setminus \{\theta_b\}$.

Now, there are two ways to motivate θ_a's focus at h_2, corresponding to the two aforementioned interpretations of limited memory. First, we can think of θ_a as a type of Ann who can see the entire

[15] For instance, in Piccione and Rubinstein's (1997) absent-minded driver's paradox, players have information sets that contain subsequent histories, the interpretation being that when they find themselves at a latter history within this information set they cannot remember what they have done at the earlier history in the same information set.

game tree, but at the same time – upon finding herself at h_2 – does not remember the utilities $(4, 0)$ that the two players would have (counterfactually) obtained if she had herself chosen L at h_0. In other words, according to this interpretation, Ann has forgotten the strategic incentives that the players had at histories outside her memory horizon. Alternatively, we can think of θ_a as a type of Ann who does remember the strategic incentives at h_0, but simply chooses to disregard them, either because it is too costly, or because she might not find it important to do so. Here we should point out that in this paper – under this second interpretation of limited memory – we do not formally model the choice of Ann's memory horizon, like for instance Dow (1991) does, and we take it as exogenously given. Nevertheless, we believe that this last part is an interesting research question for future research.

Finally, note that the two cases discussed in Section 2.3.1 are special cases of limited memory, with the players remembering everything when they reason about the entire game, and analogously the players remembering nothing when they reason only about the future.

Limited foresight. Similarly to limited memory, dynamic games with limited foresight have also been extensively studied in the literature.[16] Again, the motivation for considering the possibility of limited foresight differs depending on the context being studied, with the most prominent approach being that players are often shortsighted due to computational restrictions. Still, similarly to our discussion on limited memory, our framework allows us to think of players who are not able to reason about the future, but also of players who consciously choose to disregard some future histories. In fact, the latter scenario could arise in cases where *"looking ahead is computationally expensive and unnatural because it means reasoning about events that probably will not occur"* as Johnson et al. (2002) elaborately put it. We should repeat that while in this paper we do not study how players choose their foresight horizon, we find this question to be of interest for future research.

Formally, the idea here is that players can see the entire game, without necessarily reasoning about all future histories. Thus, we take a structure $\mathfrak{F}$, such that for every $\theta_i \in \Theta_i$, every $h \in H_i$ and every $z \in Z$ with $h \in \mathrm{Pr}(z)$, there exists a unique $h^z_{\theta_i} \in \mathrm{Fut}(h) \cap \mathrm{Pr}(z)$ with

$$F_{\theta_i}(h) = H \setminus \{h' \in \mathrm{Fut}(h) : h' \text{ strictly follows } h^z_{\theta_i} \text{ for some } z \in \mathrm{Fut}(h)\}.$$

Once again, consider the game in Fig. 1, and let θ_a be such that $F_{\theta_a}(h_0) = \{h_0, h_1\}$. This is for instance the case when Ann cannot foresee how the players will reason beyond one period ahead.

[16] In fact, limited foresight has been studied in the context of repeated games (e.g., Jehiel, 2001; Maenner, 2008), learning (e.g., Mengel, 2014), behavioral and experimental economics (e.g., Johnson et al., 2002) and computer science (e.g., Grossi and Turrini, 2012; Turrini, 2015), just to mention a few examples.

Similarly to Ex. 2, one can easily show that incomplete information about the Θ_i-types is often needed in order to model limited foresight.

Mixed cases. On top of the previous "pure" cases that we consider, our framework allows us to consider mixed reasoning structures. This is for instance the case if we want to model limited memory and limited foresight simultaneously. To the best of our knowledge, this is the first paper in the literature providing this possibility, thus also allowing us to compare the game-theoretic predictions induced by different forms of bounded rationality by means of comparative statics.

Focal histories. Besides the standard forms of bounded reasoning that we have presented above, our framework is flexible enough to accommodate other new forms of bounded reasoning that have not been systematically studied within game theory and which are often of theoretical and/or applied importance. One such case of particular interest is games with "focal histories". These are games where the choices made at certain histories have a bigger effect on the players' utilities in comparison to the choices made at other histories. Indeed, it is often the case that certain decisions carry much higher weight than others. Traditional game theory does not make any such distinction and assumes that players treat all the histories in the same way. However, it is rather natural to assume that in many strategic settings people pay attention to events that involve high stakes. For instance, in political campaigns voters tend to disregard the candidates' positions on issues that do not affect them. Also in personal relationships people often forgive/forget mistakes made by others in situations of minor importance. These informal examples suggest that our framework could be used to analyze applied problems that the existing game-theoretic literature cannot.

3. Subjective beliefs and rationality

3.1. Conditional beliefs

Using a variant of the standard framework of Battigalli and Siniscalchi (1999, 2002), we model conditional belief hierarchies by means of a type structure. Let us begin by fixing a structure $\mathfrak{F}$. Then, we consider the tuple

$$\mathcal{T}_{\mathfrak{F}} = \big((T_i)_{i\in I}, (\phi_i)_{i\in I}, (\lambda_i)_{i\in I}\big),$$

where T_i is a compact metrizable space of player i's types with typical element t_i,[17] $\phi_i : T_i \to \Theta_i$ is a surjective Borel function endowing each T_i-type with a Θ_i-type, and $\lambda_i : T_i \times H_i \to \Delta(S_{-i} \times T_{-i})$ is a

[17] Throughout the paper we often refer to elements of T_i as T_i-types whenever it is not obvious from the context and we want to distinguish them from the Θ_i-types that we introduced in the previous section.

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 $T_{-i} := \bigtimes_{j \neq i} T_j$.[18] Henceforth, we refer to the measure $\lambda_i^h(t_i)$ as t_i's conditional beliefs (or simply beliefs) at a history h. The subset

$$T_{\theta_i} := \phi_i^{-1}(\theta_i) \tag{2}$$

contains the T_i-types with $\phi_i(t_i) = \theta_i$. Observe that $\{T_{\theta_i} | \theta_i \in \Theta_i\}$ is a partition of T_i. Obviously, it is the trivial partition whenever Θ_i is a singleton. Whenever $t_i \in T_{\theta_i}$, we naturally require the conditional belief $\lambda_i^h(t_i)$ to agree with $g_i^h(\theta_i)$. This restriction is formally imposed by the following assumption.

Assumption 2. For every $i \in I$, every $h \in H_i$, every $\theta_i \in \Theta_i$, every $t_i \in T_{\theta_i}$, every $(\theta_j)_{j \neq i} \in \Theta_{-i}$, it is the case that $\lambda_i^h(t_i)\Big(S_{-i} \times \big(\bigtimes_{j \neq i} T_{\theta_j}\big)\Big) = g_i^h(\theta_i)\big((\theta_j)_{j \neq i}\big)$. $\triangleleft$

Before moving forward, observe that $S_{-i} \times (\bigtimes_{j \neq i} T_{\theta_j})$ is a Borel event in $S_{-i} \times T_{-i}$, and therefore the probability $\lambda_i^h(t_i)\big(S_{-i} \times (\bigtimes_{j \neq i} T_{\theta_j})\big)$ is well-defined. This follows directly from ϕ_j being Borel measurable for every $j \in I$.

Example 3. Recall the game in Fig. 1, together with the structure $\mathfrak{F}$ from Ex. 1. Now, consider the type structure $\mathcal{T}_{\mathfrak{F}} = (T_a, T_b, \phi_a, \phi_b, \lambda_a, \lambda_b)$, with the type spaces being $T_a = \{t_a, t'_a\}$ and $T_b = \{t_b, t'_b\}$, where $T_{\theta_a} = \{t_a\}$, $T_{\theta'_a} = \{t'_a\}$, $T_{\theta_b} = \{t_b\}$ and $T_{\theta'_b} = \{t'_b\}$. The corresponding conditional beliefs of each type are the ones shown below:

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

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. Notice that $\mathcal{T}_{\mathfrak{F}}$ satisfies Assumption 2, e.g., $t_b \in T_{\theta_b}$ and thus $\lambda_b^{h_1}(t_b)(S_a \times T_{\theta_a}) = g_b^{h_1}(\theta_b)(\theta_a) = 0.5$. $\triangleleft$

[18] 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 their Condition 3), thus implicitly assuming that the collection of conditional beliefs forms a conditional probability system, as originally defined by Rênyi (1955).

A type structure $\mathcal{T}_{\mathfrak{F}}$ 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.

Notice that by the previous construction, it is for instance the case that each t_i forms a belief at $h \in H_i$ about j's beliefs at each $h' \in H_j$, even if $t_i \in T_{\theta_i}$ and $h' \notin F_{\theta_i}(h)$. At first sight this might seem to intuitively contradict our idea of local reasoning. However, recall that our general idea is that θ_i at h can see all histories in H but only reasons about those histories that belong to $F_{\theta_i}(h)$. This distinction is formally captured by the fact that in our model θ_i's beliefs about j's beliefs at $h' \notin F_{\theta_i}(h)$ are arbitrary, i.e., $t_i \in T_{\theta_i}$ will form her beliefs without applying any form of reasoning about these histories. This rather subtle point will become clearer in Section 4, where we introduce a specific form of reasoning.

Definition 1. A type structure $\mathcal{T}_{\mathfrak{F}}$ is said to be complete if for every $i \in I$, every $\theta_i \in \Theta_i$, every $h \in H_i$, every $(\theta_j)_{j \neq i} \in \Theta_{-i}$ and every $(\mu_i^h)_{h \in H_i}$ with $\mu_i^h \in \Delta\big(S_{-i}(h) \times T_{-i}\big)$ and $\mu_i^h\big(S_{-i} \times (\bigtimes_{j \neq i} T_{\theta_j})\big) = g_i^h(\theta_i)\big((\theta_j)_{j \neq i}\big)$ for all $h \in H_i$, there exists some $t_i \in T_{\theta_i}$ such that $\lambda_i^h(t_i) = \mu_i^h$ for all $h \in H_i$.

Before moving forward, notice that the standard notion of completeness is a special case of our definition for cases where F is common knowledge. In particular, if F is commonly known, completeness postulates that 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 $h \in H_i$. To see that the standard notion of completeness agrees with our definition whenever F is commonly known, observe that in this case $T_j = T_{\theta_j}$ for the unique $\theta_j \in \Theta_j$, and therefore $\mu_i^h\big(S_{-i} \times (\bigtimes_{j \neq i} T_{\theta_j})\big) = g_i^h(\theta_i)\big((\theta_j)_{j \neq i}\big) = 1$ is trivially satisfied. Thus, our definition reduces to the usual one.

Battigalli and Siniscalchi (1999) showed the existence of a complete type structure (with commonly known F).[19] It turns out that this result can be extended to arbitrary type structures, i.e.,

[19] In fact, they proved existence of a complete type structure under Bayesian updating, but their result can be easily generalized to type structures without Bayesian updating. In a more recent paper, 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

a complete type structure $\mathcal{T}_{\mathfrak{F}}$ exists for every structure $\mathfrak{F}$, even if F is not commonly known. This claim is formally proven in Appendix A. Throughout the paper, unless explicitly stated otherwise, we work with complete type structures. 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

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

Before moving forward, let us point out that henceforth, we focus on structures $\mathfrak{F}$ with the property that each $\theta_i \in \Theta_i$ puts positive probability to a unique $\theta_{-i} \in \Theta_{-i}$ at each $h \in H_i$. Then, at some $h \in H_i$ a type t_i of player i is said to $\mathfrak{F}$-strongly believe in some event $E \subseteq S_{-i} \times T_{-i}$ whenever the following condition holds: if E does not contradict the history h, then t_i believes in E at h, i.e., formally for an arbitrary $\theta_i \in \Theta_i$,

$$SB_{\theta_i}^h(E) := \left\{ \, t_i \in T_{\theta_i} : \text{ if } (S_{-i}(h) \times T_{-\theta_i}) \cap E \neq \emptyset \text{ then } t_i \in B_i^h(E) \, \right\} \tag{3}$$

where $T_{-\theta_i} := \bigtimes_{j \neq i} T_{\theta_j}$, with $(\theta_j)_{j \neq i}$ being such that $g_i^h(\theta_i)\big((\theta_j)_{j \neq i}\big) = 1$. Notice that, unlike the standard notion of strong belief à la Battigalli and Siniscalchi (2002), here we require $(S_{-i}(h) \times T_{-\theta_i}) \cap E \neq \emptyset$ rather than $(S_{-i}(h) \times T_{-i}) \cap E \neq \emptyset$, the reason being that from θ_i's point of view the only T_j-types in the model are those in T_{θ_j}, i.e., those corresponding to Θ_j-types that θ_i deems possible in the structure $\mathfrak{F}$. The set of all types strongly believing in E are those in

$$SB_i^h(E) := \bigcup_{\theta_i \in \Theta_i} SB_{\theta_i}^h(E).$$

Obviously, if F is commonly known, then $SB_{\theta_i}^h(E) = SB_i^h(E)$, as it is the case that $T_{-\theta_i} = T_{-i}$.

Of course, it is straightforward to verify that every $t_i \in T_{\theta_i}$ strongly believes in $S_{-i}(h) \times T_{-\theta_i}$ as well as in $S_{-i}(h) \times T_{-i}$ at $h \in H_i$. Finally, we say that E is strongly believed by θ_i, whenever it is strongly believed at every $h \in H_i$, i.e., formally

$$SB_{\theta_i}(E) := \bigcap_{h \in H_i} SB_{\theta_i}^h(E),$$

updating.

and at the same time a T_i-type strongly believes in E whenever it belongs to

$$SB_i(E) := \bigcup_{\theta_i \in \Theta_i} SB_{\theta_i}(E).$$

Once again, if F is commonly known, then $SB_{\theta_i}(E) = SB_i(E)$, as it is the case that $T_{-\theta_i} = T_{-i}$.

3.2. 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

$$U_i^h(s_i, t_i) := U_i^h\big(s_i, b_i^h(t_i)\big). \tag{4}$$

In our Example 3 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

$$R_i^h := \big\{ \ (s_i, t_i) \in S_i(h) \times T_i : U_i^h(s_i, t_i) \geq U_i^h(s_i', t_i) \text{ for all } s_i' \in S_i(h) \ \big\} \ . \tag{5}$$

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 3, 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

$$\begin{aligned} R_{-i}^h \quad &:= \quad \bigtimes_{j \neq i} \ \big\{ \ (s_j, t_j) \in S_j(h) \times T_j : \ \text{if } h \in H_j \text{ then } (s_j, t_j) \in R_j^h \ \big\} \\ &= \quad \bigtimes_{j \neq i} R_j^h. \end{aligned} \tag{6}$$

denote the event that every player other than i – who is active at h – is rational at h.[20] In other words, R_{-i}^h expresses the idea that upon reaching h, all of i's active opponents at h choose a strategy

[20] 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 $j \neq i$ 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 $R^{h_1}_{-b} = S_a(h_1) \times T_a$, because Ann is not active at h_1, and therefore by our convention $R^{h_1}_a = S_a(h_1) \times T_a$. On the other hand, it is the case that $R^{h_2}_{-b} = R^{h_2}_a = \{(RA, t_a), (RA, t'_a)\}$. This is because Ann is active at h_2, and therefore $R^{h_2}_a$ is given by Eq. (5).

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

$$R^G_i := \{ \ (s_i, t_i) \in S_i \times T_i : (s_i, t_i) \in R^h_i \text{ for all } h \in H_i(s_i) \cap G \ \} \tag{7}$$

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

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

$$\begin{aligned} R^G_{-i} &:= \bigtimes_{j \neq i} \{ \ (s_j, t_j) \in S_j \times T_j : (s_j, t_j) \in R^h_j \text{ for all } h \in H_j(s_j) \cap G \ \} \\ &= \bigtimes_{j \neq i} R^G_j \end{aligned} \tag{8}$$

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 $R_{-i} := R^{H_{-i}}_{-i}$.

4. Local reasoning about the opponents' rationality

So far, we have defined a general framework which allows us to model players who reason only about some histories in the game. However, we have not specified yet *how players reason about these particular histories*. In this section, we formalize a reasoning process, based on the notion of strong belief in rationality, originally introduced by Battigalli and Siniscalchi (2002), and we introduce a solution concept that incorporates this reasoning process. This will be 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 structure $\mathfrak{F}$. 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 $\mathfrak{F}$. Moreover, our concept can be applied to make predictions in dynamic games with players exhibiting different forms of bounded rationality, such as for instance limited memory and/or limited foresight.

Let us begin by defining our notion of "*a player reasoning locally about the opponents' rationality at some histories only*". Fix an arbitrary structure $\mathfrak{F}$. Then, for an arbitrary $\theta_i \in \Theta_i$, let

$$SB^h_{\theta_i}\left(R^{F_{\theta_i}(h)}_{-i}\right) := \left\{\ t_i \in T_{\theta_i} :\ \text{if } (S_{-i}(h) \times T_{-\theta_i}) \cap R^{F_{\theta_i}(h)}_{-i} \neq \emptyset \text{ then } t_i \in B^h_i\left(R^{F_{\theta_i}(h)}_{-i}\right)\ \right\}\ . \tag{9}$$

The underlying idea is that, upon finding herself at history h, the i-agent θ_i tries to rationalize the opponents' moves at every history in $F_{\theta_i}(h)$. If them being rational at every $h' \in F_{\theta_i}(h)$ does not contradict reaching h, then θ_i will believe at h that they are indeed rational at every history $h' \in F_{\theta_i}(h)$. This type of local reasoning will be henceforth called *local strong belief in rationality with respect to* $\mathfrak{F}$ at h, or simply *strong belief in rationality* at h. Obviously, our notion of local reasoning depends on the choice of $\mathfrak{F}$. Let us illustrate this notion by means of an example.

Example 1 (cont). Recall the game in Fig. 1 and let the functions (λ_a, λ_b) be the ones from Ex. 3. Moreover, assume that F_b is such that $F_b(h_1) = \text{Fut}(h_1) = \{h_1, h_2\}$. As we have already noted above, it is the case that $R^{F_b(h_1)}_{-b} = R^{\{h_1,h_2\}}_a = \{(L, t_a), (L, t'_a), (RA, t_a), (RA, t'_a)\}$. Indeed, both $(L, t_a) \in R^{\{h_1,h_2\}}_a$ and $(L, t'_a) \in R^{\{h_1,h_2\}}_a$ hold, as $H_a(L) \cap \{h_1, h_2\} = \emptyset$. At the same time, both $(RA, t_a) \in R^{h_2}_a$ and $(RA, t'_a) \in R^{h_2}_a$ hold, and moreover h_2 is the only history in $H_a(RA) \cap \{h_1, h_2\}$, thus implying $(RA, t_a) \in R^{\{h_1,h_2\}}_a$ and $(RA, t'_a) \in R^{\{h_1,h_2\}}_a$. Hence, $SB^{h_1}_b\big(R^{F_b(h_1)}_{-b}\big) = \{t_b\}$. The reason why t'_b does not strongly believe at h_1 in Ann's rationality, is that $R^{F_b(h_1)}_{-b}$ is consistent with reaching h_1, and yet it does not receive probability 1 by $\lambda^{h_1}_b(t'_b)$, viz., $\lambda^{h_1}_b(t'_b)$ puts positive probability to (RB, t'_a). $\triangleleft$

Below, we iterate the reasoning of strong belief in rationality to obtain our solution concept of common strong belief in rationality.

4.1. Local common strong belief in rationality

Fix an arbitrary structure $\mathfrak{F}$, and take an arbitrary $\theta_i \in \Theta_i$ and an arbitrary history $h \in H_i$. Then, we define the following sequences of subsets of $T_{\theta_i} \subseteq T_i$:

$$\begin{aligned}
T^1_{\theta_i}(h) &:= SB^h_{\theta_i}\left(R^{F_{\theta_i}(h)}_{-i}\right) \\
T^2_{\theta_i}(h) &:= T^1_{\theta_i}(h) \cap SB^h_{\theta_i}\left(R^{F_{\theta_i}(h)}_{-i} \cap \left(S_{-i} \times T^1_{-\theta_i}(F_{\theta_i}(h))\right)\right) \\
&\;\;\vdots \\
T^k_{\theta_i}(h) &:= T^{k-1}_{\theta_i}(h) \cap SB^h_{\theta_i}\left(R^{F_{\theta_i}(h)}_{-i} \cap \left(S_{-i} \times T^{k-1}_{-\theta_i}(F_{\theta_i}(h))\right)\right) \\
&\;\;\vdots
\end{aligned}$$

where, for each $k > 1$,

$$T^{k-1}_{-\theta_i}\big(F_{\theta_i}(h)\big) := \bigtimes_{j \neq i} \left\{\, t_j \in T_{\theta_j} : t_j \in T^{k-1}_{\theta_j}(h') \text{ for all } h' \in F_{\theta_i}(h) \cap H_j \,\right\},$$

with $(\theta_j)_{j \neq i}$ being the unique element of Θ_{-i} receiving positive probability by θ_i at h in $\mathfrak{F}$.

Hence, $T^1_{\theta_i}(h)$ contains the types in T_{θ_i} that strongly believe at h that the opponents are rational at every $h' \in H_j \cap F_{\theta_i}(h)$. Throughout the paper, we refer to the types in

$$T^1_i(h) := \bigcup_{\theta_i \in \Theta_i} T^1_{\theta_i}(h)$$

as those satisfying *1-fold strong belief in rationality at h*.

Now, $T^2_{\theta_i}(h)$ contains those types in $T^1_{\theta_i}(h)$ that strongly believe at h that every opponent j (i) is rational at every $h' \in H_j \cap F_{\theta_i}(h)$, and (ii) strongly believe at every $h' \in H_j \cap F_{\theta_i}(h)$ that every opponent $k \neq j$ is rational at every $h'' \in H_k \cap F_{\theta_j}(h')$. The event described in (i) corresponds to $R^{F_{\theta_i}(h)}_{-i}$, while the event described in (ii) corresponds to

$$S_{-i} \times T^1_{-\theta_i}\big(F_{\theta_i}(h)\big) = \bigtimes_{j \neq i} \left\{\, (s_j, t_j) \in S_j \times T_{\theta_j} : t_j \in SB^{h'}_{\theta_j}\big(R^{F_{\theta_j}(h')}_{-\theta_j}\big) \text{ for all } h' \in F_{\theta_i}(h) \cap H_j \,\right\}$$

in the second equation of the sequence above. The reason for explicitly requiring every type in $T^2_{\theta_i}(h)$ to belong to $T^1_{\theta_i}(h)$ is that the strong belief operator is not monotonic, thus implying that $SB^h_{\theta_i}(E \cap F)$ is not necessarily equal to $SB^h_{\theta_i}(E) \cap SB^h_{\theta_i}(F)$.[21] Therefore $SB^h_{\theta_i}\big(R^{F_{\theta_i}(h)}_{-i}\big)$ does not follow directly from $SB^h_{\theta_i}\big(R^{F_{\theta_i}(h)}_{-i} \cap \big(S_{-i} \times T^1_{-\theta_i}(F_{\theta_i}(h))\big)\big)$. Throughout the paper, we refer to the types in

$$T^2_i(h) := \bigcup_{\theta_i \in \Theta_i} T^2_{\theta_i}(h)$$

[21] 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.

as those satisfying *up to 2-fold strong belief in rationality at* h. The reason we add the term "up to" is that, by construction, $T_i^2(h) \subseteq T_i^1(h)$, as we have already discussed above.

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

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

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

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

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

4.2. Local iterated conditional dominance procedure

In this section we introduce a (finite) procedure which, for every player $i \in I$, for every i-agent $\theta_i \in \Theta_i$ and every history $h \in H_i$, iteratively eliminates (at each round) strategies from $S_i(h)$ and first order conditional beliefs from $\Delta\big(S_{-i}(h)\big)$.[22] 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 in Perea (2014).[23] Before formally defining our procedure, let 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 $i \in I$, for each $\theta_i \in \Theta_i$ and each history $h \in H_i$.

Decision problem. A decision problem for $\theta_i \in \Theta_i$ at a history $h \in H_i$ is a tuple $\big(B_{\theta_i}(h), D_{\theta_i}(h)\big)$, with $B_{\theta_i}(h) \subseteq S_{-i}(h)$ and $D_{\theta_i}(h) \subseteq S_i(h)$. Intuitively, $B_{\theta_i}(h)$ can be seen as the subset of the opponents' strategies that θ_i could deem possible at h. At this point, we should already make clear that the link between $B_{\theta_i}(h)$ and what θ_i could deem possible at h is only an informal one. The

[22] In fact, as we are going to see later in the paper, it might be the case that different strategies and/or conditional beliefs are eliminated for different Θ_i-types at the same history.

[23] Later in the paper, we discuss the relationship of our procedure with the iterated conditional dominance procedure and with the backward dominance procedure.

actual relationship between the two will become apparent later on in the paper. Thus, for the time being, $B_{\theta_i}(h)$ and $D_{\theta_i}(h)$ will be merely treated as auxiliary tools, without a concrete meaning.

A strategy $s_i \in D_{\theta_i}(h)$ is said to be rational in the decision problem $\big(B_{\theta_i}(h), D_{\theta_i}(h)\big)$ whenever there exists a probability measure $\beta_i^h \in \Delta\big(B_{\theta_i}(h)\big)$ such that $U_i^h(s_i, \beta_i^h) \geq U_i^h(s_i', \beta_i^h)$ for all $s_i' \in D_{\theta_i}(h)$. 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 structure $\mathfrak{F}$, our procedure will be defined by means of a (weakly) decreasing sequence $\big(B_{\theta_i}^k(h), D_{\theta_i}^k(h)\big)_{k\geq 0}$ of decision problems for each $i \in I$, each $\theta_i \in \Theta_i$ 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

$$\begin{aligned} B_{\theta_i}^0(h) &:= S_{-i}(h) \\ D_{\theta_i}^0(h) &:= S_i(h). \end{aligned}$$

Obviously, this initial step does not depend on the choice of $\mathfrak{F}$.

Inductive step of the procedure. Now, fix some $k > 0$ and suppose that for each $i \in I$, each $\theta_i \in \Theta_i$ and each $h \in H_i$ we have undertaken the $(k-1)$-th step of our procedure, thus having obtained $\big(B_{\theta_i}^{k-1}(h), D_{\theta_i}^{k-1}(h)\big)$. Then, for an arbitrary $h \in H_i$, define $\big(B_{\theta_i}^k(h), D_{\theta_i}^k(h)\big)$ by

$$B_{\theta_i}^k(h) := \begin{cases} C_{\theta_i}^{k-1}(h) & \text{if } C_{\theta_i}^{k-1}(h) \neq \emptyset \\ B_{\theta_i}^{k-1}(h) & \text{if } C_{\theta_i}^{k-1}(h) = \emptyset \end{cases} \tag{12}$$

$$D_{\theta_i}^k(h) := \big\{\, s_i \in D_{\theta_i}^{k-1}(h) : s_i \text{ is rational in } \big(B_{\theta_i}^k(h), D_{\theta_i}^{k-1}(h)\big) \,\big\}, \tag{13}$$

where

$$C_{\theta_i}^{k-1}(h) := \bigtimes_{j\neq i} \big\{\, s_j \in S_j(h) : s_j \in D_{\theta_j}^{k-1}(h') \text{ for all } h' \in H_j(s_j) \cap F_{\theta_i}(h) \,\big\}, \tag{14}$$

with $(\theta_j)_{j\neq i}$ being the unique element of Θ_{-i} receiving positive probability by θ_i at h in $\mathfrak{F}$.

The underlying idea behind our procedure is as follows: First, for each $\theta_i \in \Theta_i$ and each $h \in H_i$, we compute $C_{\theta_i}^{k-1}(h)$ which contains all strategy combinations of i's opponents which (i) are consistent with reaching h, and (ii) have not been eliminated from $D_{\theta_j}^{k-1}(h')$ at any $h' \in F_{\theta_i}(h)$ and for any j-agent θ_j who is active at h' and is deemed possible by θ_i at h in the structure $\mathfrak{F}$. Notice that in principle $C_{\theta_i}^{k-1}(h)$ might be empty. To see this, consider for instance the game in Fig. 1 with the commonly known F being such that $F_b(h_1) = \{h_0, h_2\}$, and assume that $D_a^{k-1}(h_0) = \{L\}$. Then, clearly it is the case that $S_a(h_1) \cap D_a^{k-1}(h_0) = \emptyset$, thus implying that $C_b^{k-1}(h_1) = \emptyset$.

Having defined $C_{\theta_i}^{k-1}(h)$, we can now proceed to the k-th step of our procedure, by first defining $B_{\theta_i}^{k}(h)$. In particular, a strategy combination $s_{-i} = (s_j)_{j \neq i}$ is eliminated from $B_{\theta_i}^{k-1}(h)$ if and only if (i) there exists some history $h' \in F_{\theta_i}(h) \cap H_j(s_j)$ such that s_j has been eliminated from $D_{\theta_j}^{k-1}(h')$ for the Θ_j-type that θ_i deems possible at h, and also (ii) there exists another strategy combination $s'_{-i} = (s'_j)_{j \neq i} \in B_{\theta_i}^{k-1}(h)$ such that for every $j \neq i$ and every $h' \in F_{\theta_i}(h) \cap H_j(s_j)$ it is the case that $s'_j \in D_{\theta_j}^{k-1}(h')$ for the same Θ_j-type θ_j that θ_i deems possible at h, i.e., not all strategy combinations are eliminated from $B_{\theta_i}^{k-1}(h)$.

Now, once we have obtained $B_{\theta_i}^{k}(h)$, we can define the decision problem $\big(B_{\theta_i}^{k}(h), D_{\theta_i}^{k-1}(h)\big)$, and we eliminate from $D_{\theta_i}^{k-1}(h)$ the strategies that are not rational in this decision problem. Then, it follows from Pearce (1984, Lem. 3) that a strategy is eliminated from $D_{\theta_i}^{k-1}(h)$ if and only if it is strictly dominated by a mixed strategy within this decision problem.

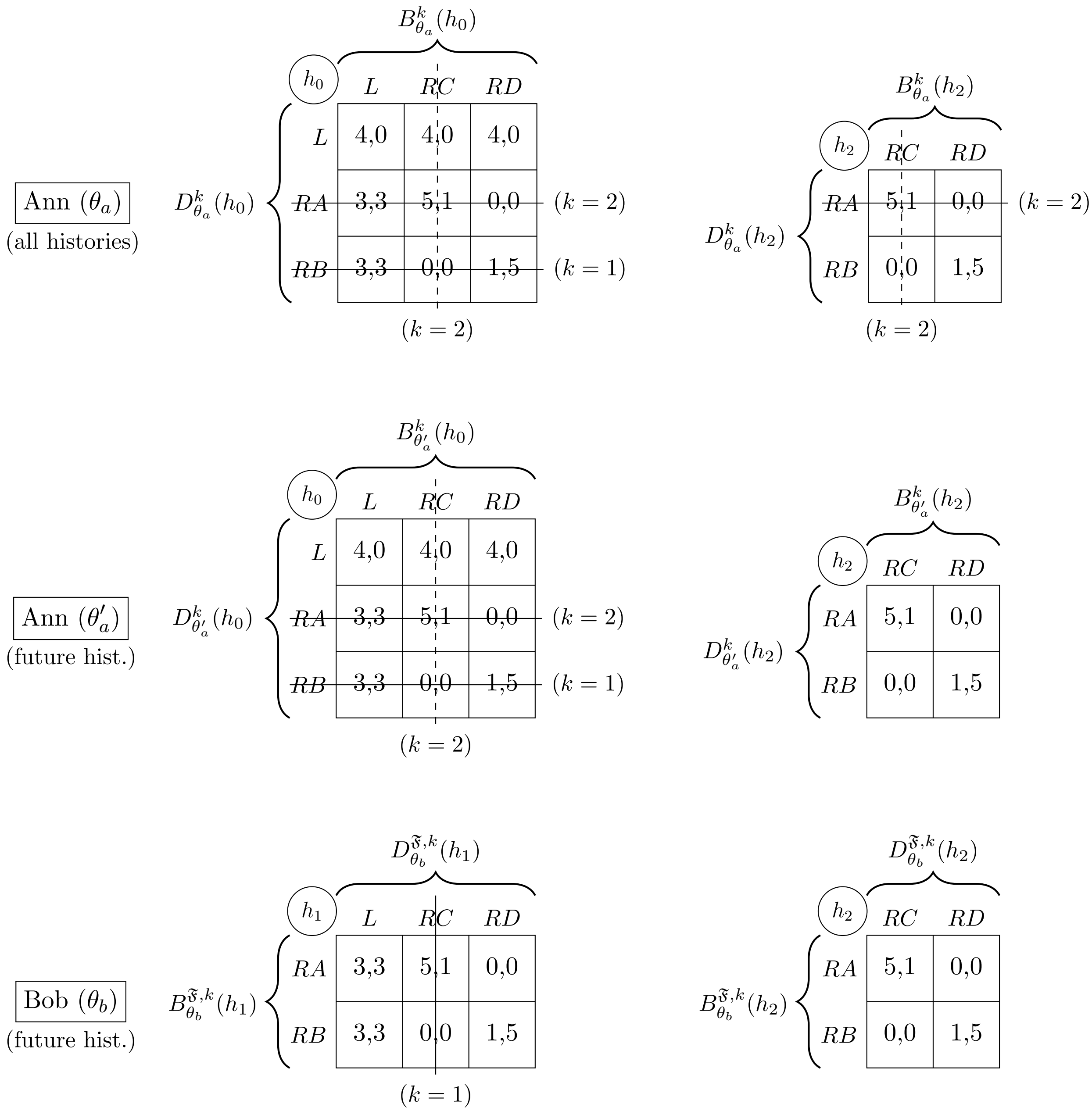
This elimination procedure is called *iterated conditional dominance procedure ($\mathfrak{F}$-ICDP)*. Obviously, since we consider only finite dynamic games and structures with finitely many Θ_i-types for each player, $\mathfrak{F}$-ICDP converges after finitely many steps. That it, there exists some $K \geq 0$ such that for each $k \geq K$, for every $i \in I$, every $\theta_i \in \Theta_i$ and every $h \in H_i$, it is the case that $\big(B_{\theta_i}^{k}(h), D_{\theta_i}^{k}(h)\big) = \big(B_{\theta_i}^{K}(h), D_{\theta_i}^{K}(h)\big)$. Then, we write $\big(B_{\theta_i}^{\mathfrak{F}}(h), D_{\theta_i}^{\mathfrak{F}}(h)\big) = \big(B_{\theta_i}^{K}(h), D_{\theta_i}^{K}(h)\big)$. We say that a *strategy s_i survives the iterated conditional dominance procedure* for some θ_i if it is the case that $s_i \in D_{\theta_i}^{\mathfrak{F}}(h)$ for all $h \in H_i(s_i)$. Below, we illustrate the $\mathfrak{F}$-ICDP with a simple example.

Example 4. Recall the example of Fig. 1, and let Ann always focus on all histories, while Bob focuses only on future histories. Furthermore, it is commonly believed that both players focus on the future histories, i.e., formally, the structure $\mathfrak{F}$ is such that $\Theta_a = \{\theta_a, \theta'_a\}$ and $\Theta_b = \{\theta_b\}$ with

$$\begin{array}{llll}
F_{\theta_a}(h_0) = H, & F_{\theta_a}(h_2) = H, & g_a^{h_0}(\theta_a) = (1 \otimes \theta_b), & g_a^{h_2}(\theta_a) = (1 \otimes \theta_b) \\
F_{\theta'_a}(h_0) = H, & F_{\theta'_a}(h_2) = \{h_2\}, & g_a^{h_0}(\theta'_a) = (1 \otimes \theta_b), & g_a^{h_2}(\theta'_a) = (1 \otimes \theta_b) \\
F_{\theta_b}(h_1) = \{h_1, h_2\}, & F_{\theta_b}(h_2) = \{h_2\}, & g_b^{h_1}(\theta_b) = (1 \otimes \theta'_a), & g_b^{h_2}(\theta_b) = (1 \otimes \theta'_a)
\end{array}$$

Let us now depict each decision problem $\big(B_{\theta_i}^{k}(h), D_{\theta_i}^{k}(h)\big)$ with a normal form game. The steps of the $\mathfrak{F}$-ICDP are represented by the lines that cross out the corresponding strategies. Eliminations from $B_{\theta_i}^{k}(h)$ are represented by dashed lines, whereas eliminations from $D_{\theta_i}^{k}(h)$ 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 $B_{\theta_i}^{0}(h)$ for any $\theta_i \in \Theta_i$, any $i \in I$ and any $h \in H_i$. Then, RB is eliminated from $D_{\theta_a}^{0}(h_0)$ and from $D_{\theta'_a}^{0}(h_0)$ because it is strictly



dominated by L at h_0, and likewise RC eliminated from $D^0_{\theta_b}(h_1)$ because it is strictly dominated by L at h_1. Hence, we obtain $D^1_{\theta_a}(h_0) = D^1_{\theta'_a}(h_0) = \{L, RA\}$ and $D^1_{\theta_b}(h_1) = \{L, RD\}$. Furthermore, no strategy is eliminated for any type at h_2, i.e., it is the case that $D^1_{\theta_a}(h_2) = D^1_{\theta'_a}(h_2) = \{RA, RB\}$ and $D^1_{\theta_b}(h_2) = \{RC, RD\}$.

At the second step ($k = 2$), Bob's strategy RC is eliminated both from $B^1_{\theta_a}(h_0)$ and $B^1_{\theta'_a}(h_0)$. This is because at h_0 both Θ_a-types of Ann (i) reason about h_1, and (ii) deem θ_b possible. On the other hand, RC is eliminated from $B^1_{\theta_a}(h_2)$ but not from $B^1_{\theta'_a}(h_2)$, the reason being that h_1 belongs to $F_{\theta_a}(h_2)$ but not to $F_{\theta'_a}(h_2)$. Thus, we obtain $B^2_{\theta_a}(h_0) = B^2_{\theta'_a}(h_0) = \{L, RD\}$, $B^2_{\theta_a}(h_2) = \{RD\}$ and $B^2_{\theta'_a}(h_2) = \{RC, RD\}$. Then, RA is eliminated from $D^1_{\theta_a}(h_0)$ and from $D^1_{\theta'_a}(h_0)$, as it is strictly dominated by L at h_0 for both Θ_a-types. Furthermore, RA is also eliminated from $D^1_{\theta_a}(h_2)$ but not from $D^1_{\theta'_a}(h_2)$, because it is strictly dominated by RB at h_2 in the decision problem $\big(B^2_{\theta_a}(h_2), D^1_{\theta_a}(h_2)\big)$ but not in the decision problem $\big(B^2_{\theta'_a}(h_2), D^2_{\theta'_a}(h_2)\big)$. Now, switching our attention to Bob's unique Θ_b-type θ_b, RB is not eliminated either from $B^1_{\theta_b}(h_1)$ or from $B^1_{\theta_b}(h_2)$. This is because h_0 does not belong to $F_{\theta_b}(h_1)$ or to $F_{\theta_b}(h_2)$. Also, RA is not eliminated from either $B^2_{\theta_b}(h_1)$ or $B^2_{\theta_b}(h_2)$, because θ_b believes that Ann is of the Θ_a-type θ'_a.

Notice that the procedure stops after two rounds of elimination. The strategies of Ann that survive $\mathfrak{F}$-ICDP are L for θ_a and for θ'_a, while for Bob's unique Θ_b-type θ_b the strategies that survive the procedure are L and RD. Indeed, observe for instance that $L \in D^{\mathfrak{F}}_{\theta_b}(h)$ for every $h \in H_b(L) = \{h_1\}$ and also $RD \in D^{\mathfrak{F}}_{\theta_b}(h)$ for every $h \in H_b(RD) = \{h_1, h_2\}$.

At this point, we should also point out that the procedure yields, not only the strategy profiles that survive for each Θ_i-type, but also the conditional beliefs at each history, e.g., according to the procedure, the only belief that θ_a can have at h_2 is to put probability 1 to Bob playing according to the strategy RD. Observe that this differs from θ'_a's conditional beliefs at the same history h_2. This is not surprising, given that θ_a and θ'_a reason about different histories while being at h_2. Below, we further elaborate on the fact that the procedure simultaneously induces strategies and conditional beliefs for each Θ_i-type of 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 $i \in I$, each $\theta_i \in \Theta_i$ and each $h \in H_i$, first we eliminate opponents' strategy combinations from $B^k_{\theta_i}(h)$, and then we eliminate strategies from $D^k_{\theta_i}(h)$. Note that these two types of elimination are conceptually very different. Let us for the time being focus on $B^k_{\theta_i}(h)$.

Eliminating a strategy combination $s_{-i} \in S_{-i}(h)$ from $B^k_{\theta_i}(h)$ can be thought of as eliminating all of θ_i's first order conditional beliefs at h that put positive probability to s_{-i}. Consequently, this elimination can be interpreted as a restriction imposed on θ_i's types, viz., eliminating s_{-i} from $B^k_{\theta_i}(h)$ essentially means that we are ruling out all types $t_i \in T_{\theta_i}$ with the property that $\mathrm{marg}_{S_{-i}} \lambda^h_i(t_i)(\{s_{-i}\}) > 0$. But then recall that this is exactly what $\mathfrak{F}$-CSBR does, i.e., it recursively imposes restrictions on θ_i's types. In the next section we show that there is indeed a very tight relationship between eliminating opponents' strategies from $B^k_{\theta_i}(h)$ and eliminating types from $T^{k-1}_{\theta_i}(h)$. Thus, it becomes clear why earlier in this section we stated that the strategy combinations in $B^k_{\theta_i}(h)$ can be thought as those that θ_i could deem possible at h after k rounds of reasoning.

4.3. Characterization results

As we have already mentioned in the previous section, there is a very tight relationship between the process of eliminating types from $T^{k-1}_{\theta_i}(h)$ and the process of eliminating opponents' strategy profiles from $B^k_{\theta_i}(h)$. The following result makes this relationship formal.

Theorem 1. *Fix a structure $\mathfrak{F}$ and consider a complete type structure $\mathcal{T}_{\mathfrak{F}}$. Then, for every player $i \in I$, every $\theta_i \in \Theta_i$, every history $h \in H_i$ and every $k > 0$, the following hold:*

(i) If $t_i \in T^{k-1}_{\theta_i}(h)$ then there exists some $\beta^h_i \in \Delta\big(B^k_{\theta_i}(h)\big)$ with $b^h_i(t_i) = \beta^h_i$.

(ii) If $\beta^h_i \in \Delta\big(B^k_{\theta_i}(h)\big)$ then there exists some $t_i \in T^{k-1}_{\theta_i}(h)$ with $b^h_i(t_i) = \beta^h_i$.

For instance, in the context of Ex. 4 the previous result implies that, for every type $t_a \in T^1_{\theta_a}(h_2)$ it is the case that $b^{h_0}_a(t_a)$ puts probability 0 to RC. This is because $B^2_{\theta_a}(h_2) = \{RD\}$. Still, we should stress that part (ii) in the theorem above does not say that every t_i with $b^h_i(t_i) \in \Delta\big(B^k_{\theta_i}(h)\big)$ belongs to $T^{k-1}_{\theta_i}(h)$. To see this, consider a type $t_i \in T_{\theta_i}$ which at $h \in H_i$ puts probability 1 to a strategy-type combination $(s_{-i}, t_{-i}) \in B^k_{\theta_i}(h) \times \big(T^{k-3}_{-\theta_i}(h') \setminus T^{k-2}_{-\theta_i}(h')\big)$ where $h' \in F_{\theta_i}(h)$, implying that $b^h_i(t_i) \in \Delta\big(B^k_{\theta_i}(h)\big)$ and also $t_i \notin T^{k-1}_{\theta_i}(h)$. Notice that such a type exists whenever the type structure is complete.

With this result at hand, we can then characterize the strategies that can be rationally played under $\mathfrak{F}$-CSBR, by means of the $\mathfrak{F}$-ICDP.

Theorem 2. *Fix a structure $\mathfrak{F}$ and consider a complete type structure $\mathcal{T}_{\mathfrak{F}}$. Then, for each player $i \in I$ and each $\theta_i \in \Theta_i$, it is the case that $s_i \in \mathrm{Proj}_{S_i}\big(R_i \cap (S_i \times T^{\mathfrak{F}}_{\theta_i})\big)$ if and only if $s_i \in D^{\mathfrak{F}}_{\theta_i}(h)$ for all $h \in H_i(s_i)$.*

The previous result formally states that a strategy can be rationally played (by a Θ_i-type θ_i) under $\mathfrak{F}$-CSBR if and only it survives the $\mathfrak{F}$-ICDP (for θ_i). For instance, in the context of Ex. 4, the only strategy that can be rationally played by θ_a under $\mathfrak{F}$-CSBR is L, as this is the only strategy surviving the $\mathfrak{F}$-ICDP. Likewise, the strategies that can be rationally played by θ_b under $\mathfrak{F}$-CSBR are L and RD, as both of them survive the $\mathfrak{F}$-ICDP.

5. Special cases of local reasoning

In this section we present some special cases of $\mathfrak{F}$. As we have already mentioned earlier in the paper, in some of these cases with $F \in \mathcal{F}$ being commonly known, $\mathfrak{F}$-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). Below, we explicitly show the equivalence to these standard cases.

5.1. 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.[24]

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:

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

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 consider a structure $\mathfrak{F}$ such that $(F_i)_{i \in I}$ is commonly known, with $F_i(h) = H$ for all $h \in H_i$ and all $i \in I$. Then, we ask whether there is a formal relationship between Battigalli and Siniscalchi's (2002) CSBR on the one hand and our $\mathfrak{F}$-CSBR on the other. As it turns out the two notions are equivalent, as shown below.

[24] 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; Reny, 1992; McLennan, 1985; Hillas, 1994).

Proposition 1. *Let the structure $\mathfrak{F}$ be such that $(F_i)_{i\in I}$ is commonly known with $F_i(h) = H$ for all $i \in I$ and all $h \in H_i$, and consider an arbitrary type structure $\mathcal{T}_{\mathfrak{F}}$. Then, for every player $i \in I$ and every $k > 0$, it is the case that $SB_i^k = \bigcap_{h\in H_i} T_i^k(h)$.*

Two immediate conclusions follow from the previous result. First, a type satisfies CSBR if and only it satisfies $\mathfrak{F}$-CSBR. Then, it naturally follows that a strategy can be rationally played under CSBR if and only if it can be rationally played under $\mathfrak{F}$-CSBR. This is formally stated in the following corollary. The proof trivially follows from the definition of rationality.

Corollary 1. *Let the structure $\mathfrak{F}$ be such that $(F_i)_{i\in I}$ is commonly known with $F_i(h) = H$ for all $i \in I$ and all $h \in H_i$, and consider an arbitrary type structure $\mathcal{T}_{\mathfrak{F}}$. Then, for every player $i \in I$, it is the case that $\mathrm{Proj}_{S_i}\big(R_i \cap (S_i \times CSB_i)\big) = \mathrm{Proj}_{S_i}\big(R_i \cap (S_i \times T_i^{\mathfrak{F}})\big)$.*

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 a strategy survives k steps of our $\mathfrak{F}$-ICDP if and only if it survives k steps of Shimoji and Watson's (1998) ICDP. In this sense, ICDP is a special case of $\mathfrak{F}$-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 Pearce's (1984) Extensive Form Rationalizability (EFR) (Battigalli and Siniscalchi, 2002). 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 $\mathfrak{F}$-CSBR also yield an EFBRS.

5.2. 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.[25]

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

[25] 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, 2015).

(BR) (Penta, 2015).[26] 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).[27] 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:

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

where $FB_{-i}^{k-1} := \bigtimes_{j \neq i} FB_j^{k-1}$ for each $k > 1$. We say that FB_i^k contains the types that satisfy *up to k-fold belief in future rationality.*[28] 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 $s_i \in \mathrm{Proj}_{S_i}\big(R_i \cap (S_i \times CFB_i)\big)$.

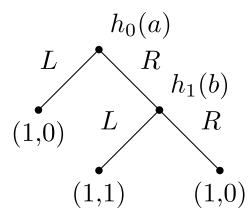
[26]Concepts like subgame perfect equilibrium (Selten, 1965) or sequential equilibrium (Kreps and Wilson, 1982) impose additional equilibrium conditions, whereas the standard backward induction procedure is well-defined only for perfect information extensive-form games without relevant ties.

[27]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.

[28]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).

The previous idea – of backward induction postulating that players disregard the future – is formally captured by the assumption that $F_i(h) = \mathrm{Fut}(h)$ 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 strong belief in rationality on the other. First, let us point out that whenever $\mathfrak{F}$ is such that $F_i(h) = \mathrm{Fut}(h)$, it is by definition the case that strong belief is directly reduced to standard belief. The reason is that rationality in $F_i(h)$ will never be contradicted by what i has already observed, as $F_i(h)$ does not contain any past history, i.e., formally speaking, it is the case that $(S_{-i}(h) \times T_{-i}) \cap R_{-i}^{F_i(h)} \neq \emptyset$. Thus, it is not surprising that the two notions are equivalent in terms of the strategy profiles they 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 $\mathfrak{F}$-CSBR.

Example 5. Consider the following dynamic game between Ann and Bob, and let (F_a, F_b) be commonly known, with $F_i(h) = \mathrm{Fut}(h)$ for each $i \in \{a, b\}$ and each $h \in H_i$, i.e., $F_a(h_0) = \{h_0, h_1\}$ and $F_b(h_1) = \{h_1\}$.



Now, consider the type structure $\mathcal{T}_{\mathfrak{F}}$ 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.

Now, switching our attention to $\mathfrak{F}$-CSBR, observe that again the only type of Ann satisfying 1-fold strong belief in rationality at h_0 is t'_a, viz., $T_a^1(h_0) = \{t'_a\}$. But, then t_b does satisfy up to 2-fold strong belief in rationality at h_1. This is because $F_b(h_1) = \{h_1\}$, and therefore $T_b^k(h_1) = T_b$ for all $k > 0$. $\triangleleft$

The reason for the previously illustrated divergence between FB_b^2 and $\bigcap_{h \in H_b} T_b^2(h)$ 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 strong belief in rationality, he 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 $\mathfrak{F}$-CSBR with $F_i(h) = \mathrm{Fut}(h)$ 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 $\mathfrak{F}$-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 $\mathfrak{F}$-CSBR if and only if it can be rationally played under $CBFR$.

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

The proof of the result follows almost directly from Lemma B3 in Appendix B, which formally proves that BDP and $\mathfrak{F}$-ICDP are essentially equivalent.

Finally, notice that while $\mathfrak{F}$-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 5. In particular, observe that, given the type structure that we assume, Rationality and CBFR yields an empty set of predictions, whereas Rationality and $\mathfrak{F}$-CSBR induces a non-empty prediction, viz., $\mathrm{Proj}_{S_b}\big(R_b \cap (S_b \times T_b^{\mathfrak{F}})\big) = \{L\}$, while $\mathrm{Proj}_{S_b}\big(R_b \cap (S_b \times CFB_b)\big) = \emptyset$.

6. Discussion

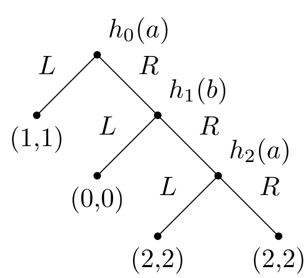
6.1. Infinite structures

Throughout the paper we have focused exclusively on finite structures $\mathfrak{F}$. The main reason for doing so is for simplicity. In our view, almost all interesting cases of local reasoning can be already modelled within a finite structure and therefore we do not believe that allowing for arbitrary structures would provide any additional insight. Still, it is natural to ask whether our results also hold in the general case. Our conjecture is that they do, but one should first impose some additional topological/measure-theoretic structure. For instance, if we take some $\mathfrak{F}$ with Θ_i being an arbitrary compact metrizable space, we may need to impose additional restrictions in order to construct the complete type structure $\mathcal{T}_{\mathfrak{F}}$. The reason is that in our analysis different Θ_i-types of player i are often treated as if they were different players. Finally, on a more practical issue, one of the major advantages of our $\mathfrak{F}$-ICDP is its tractability. This would not be the case anymore if $\mathfrak{F}$ was infinite, and therefore it would become challenging to compute the predictions of the model.

6.2. Relationship to unawareness

As we have already mentioned in the paper, our framework differs from the one concerning dynamic games with unawareness, not only in its technical aspects but also conceptually. In fact, observe that the underlying idea behind all models of unawareness is that players cannot even "see" some parts of the game, such as certain moves or even entire histories (e.g., Feinberg, 2012; Heifetz et al., 2013; Halpern and Rêgo, 2014). As a consequence, players do not reason about these parts of the game, similarly to what happens in our framework. However, the difference between the two is that in games with unawareness players do not even form beliefs about these parts of the game, whereas in our model the players do form beliefs, but these beliefs are completely arbitrary.

This distinction is sometimes crucial, not only conceptually, but also for our predictions. To see this consider the following simple game, supposing that at h_0 Ann does not reason about h_2, thus



exhibiting limited foresight. Still Ann can figure out that Bob will choose R at h_1, even without thinking about how Bob (at h_1) will reason about her behavior at h_2. As a consequence, our concept predicts that she will choose R at h_0. On the other hand, if she was unaware at h_0 of the existence of R at h_1, she would choose L at h_0.

A. Complete type structures

In this section we construct a canonical type structure $\mathcal{T}_{\mathfrak{F}}$ for any structure $\mathfrak{F}$. This type structure will be complete. For notation simplicity, we prove our claim for two players, but our result can be directly generalized to any finite number of players.

We begin by fixing some $\mathfrak{F} = \big((\Theta_i)_{i\in I}, (f_i)_{i\in I}, (g_i)_{i\in I}\big)$, and for each $h \in H_i$ we consider the following sequence of spaces:

$$\begin{aligned}
\Omega_i^0(h) &:= \Theta_i \times S_j(h) \\
\Omega_i^1(h) &:= \Omega_i^0(h) \times \Big(\bigtimes_{h' \in H_j} \Delta\big(\Omega_j^0(h')\big)\Big) \\
&\vdots \\
\Omega_i^{k+1}(h) &:= \Omega_i^k(h) \times \Big(\bigtimes_{h' \in H_j} \Delta\big(\Omega_j^k(h')\big)\Big) \\
&\vdots
\end{aligned}$$

Obviously, $\Omega_i^k(h)$ is a compact metrizable space for every $k \geq 0$. Then, we define the product spaces $\tilde{T}_i^0(h) := \bigtimes_{k\geq 0} \Delta\big(\Omega_i^k(h)\big)$ and $\tilde{T}_i^0 := \bigtimes_{h\in H_i} \tilde{T}_i^0(h)$. For notation simplicity, let us denote the typical element of $\tilde{T}_i^0$ by t_i. Hence, each $t_i \in \tilde{T}_i^0$ is essentially an abbreviation for $\big(\mu_{t_i}^1(h), \mu_{t_i}^2(h), \dots\big)_{h\in H_i}$, with $\mu_{t_i}^k(h) \in \Delta\big(\Omega_i^{k-1}(h)\big)$ standing for the corresponding coordinate of t_i. As usual, we impose the standard coherency condition, thus restricting attention to collections of conditional beliefs in

$$\tilde{T}_i^1 := \big\{ \ t_i \in \tilde{T}_i^0 : \mathrm{marg}_{\Omega_i^k(h)}\, \mu_{t_i}^{k+2}(h) = \mu_{t_i}^{k+1}(h) \text{ for all } k \geq 0 \text{ and all } h \in H_i \ \big\}$$

Then, it follows from earlier works of Brandenburger and Dekel (1993) and Battigalli and Siniscalchi (1999) that there exists a homeomorphism $\tilde{\pi}_i : \tilde{T}_i^1 \to \bigtimes_{h\in H_i} \Delta\big(\Theta_i \times S_j(h) \times \tilde{T}_j^0\big)$, with $\tilde{\pi}_i^h(t_i) := \mathrm{Proj}_{\Delta(\Theta_i \times S_j(h) \times \tilde{T}_j^0)}\, \tilde{\pi}_i(t_i)$. In fact, this is a direct consequence of Kolmogorov extension theorem. Now, for each $k > 1$, we recursively define

$$\tilde{T}_i^k := \big\{ \ t_i \in \tilde{T}_i^{k-1} : \tilde{\pi}_i^h(t_i)\big(\Theta_i \times S_j(h) \times \tilde{T}_j^{k-1}\big) = 1 \text{ for all } h \in H_i \ \big\}$$

and we let $\tilde{T}_i := \bigcap_{k\geq 0} \tilde{T}_i^k$ be the set of conditional belief hierarchies that satisfy coherency and common certainty in coherency. Then, again from Brandenburger and Dekel (1993) and Battigalli and Siniscalchi (1999) it follows that there exists a homeomorphism $\pi_i : \tilde{T}_i \to \bigtimes_{h\in H_i} \Delta\big(\Theta_i \times S_j(h) \times \tilde{T}_j\big)$, once again with $\pi_i^h(t_i) := \mathrm{Proj}_{\Delta(\Theta_i \times S_j(h) \times \tilde{T}_j)}\, \pi_i(t_i)$. Note that $\tilde{T}_i$ is a compact metrizable space.

Now, for an arbitrary $\theta_i \in \Theta_i$, define

$$\hat{T}_{\theta_i}^1 := \big\{ \ t_i \in \tilde{T}_i : \pi_i^h(t_i)\big(\{\theta_i\} \times S_j(h) \times \tilde{T}_j\big) = 1 \text{ for all } h \in H_i \ \big\}$$

and let $\hat{T}_i^1 := \bigcup_{\theta_i \in \Theta_i} \hat{T}_{\theta_i}^1$. Observe that $\{\theta_i\}$ is closed in Θ_i, and therefore $\hat{T}_{\theta_i}^1$ is closed in $\tilde{T}_i$ (Aliprantis and Border, 1994, Cor. 15.6). Thus, $\hat{T}_i^1$ is also closed in $\tilde{T}_i$, as it is the finite union of closed subsets. In fact, $\hat{T}_i^1$ is also open, as the complement of the closed subset $\bigcup_{\theta_i' \in \Theta_i \setminus \{\theta_i\}} \hat{T}_{\theta_i'}^1$. Then, for every $\theta_i \in \Theta_i$ and every $k > 1$, we recursively define

$$\hat{T}_{\theta_i}^k := \big\{ \ t_i \in \hat{T}_{\theta_i}^{k-1} : \pi_i^h(t_i)\big(\{\theta_i\} \times S_j(h) \times \hat{T}_j^{k-1}\big) = 1 \ \big\}$$

and we let $\hat{T}_{\theta_i} := \bigcap_{k\geq 0} \hat{T}^k_{\theta_i}$ and moreover $\hat{T}_i := \bigcup_{\theta_i\in\Theta_i} \hat{T}_{\theta_i}$. Using a similar argument as above, it follows from Aliprantis and Border (1994, Cor. 15.6) that $\hat{T}^k_{\theta_i}$ is closed. Hence, $\hat{T}_{\theta_i}$ is also closed as the intersection of closed subsets, and therefore so is $\hat{T}_i$ as the finite union of closed sets.

Now, for an arbitrary $\theta_i \in \Theta_i$, define

$$T^1_{\theta_i} := \left\{ t_i \in \hat{T}_{\theta_i} : \pi^h_i(t_i)\big(\{\theta_i\} \times S_j(h) \times \hat{T}_{\theta_j}\big) = g^h_i(\theta_i)(\theta_j) \text{ for all } \theta_j \in \Theta_j \text{ and all } h \in H_i \right\} .$$

Observe that $\hat{T}_{\theta_j}$ is clopen, and therefore both $\{t_i \in \hat{T}_{\theta_i} : \pi^h_i(t_i)(\{\theta_i\} \times S_j(h) \times \hat{T}_{\theta_j}) \geq g^h_i(\theta_i)(\theta_j)\}$ and $\{t_i \in \hat{T}_{\theta_i} : \pi^h_i(t_i)(\{\theta_i\} \times S_j(h) \times \hat{T}_{\theta_j}) \leq g^h_i(\theta_i)(\theta_j)\}$ are closed (Aliprantis and Border, 1994, Cor. 15.6), thus implying that so is $T^1_{\theta_i}$. Then, for every $\theta_i \in \Theta_i$ and every $k > 1$, we recursively define

$$T^k_{\theta_i} := \left\{ t_i \in T^{k-1}_{\theta_i} : \pi^h_i(t_i)\big(\{\theta_i\} \times S_j(h) \times T^{k-1}_j\big) = 1 \right\}$$

and let $T_{\theta_i} := \bigcap_{k\geq 0} T^k_{\theta_i}$ and moreover $T_i := \bigcup_{\theta_i\in\Theta_i} T_{\theta_i}$. Following the same steps as above, we show that T_{θ_i} and T_i are clopen and therefore compact metrizable subspaces.

Following Brandenburger and Dekel (1993) and Battigalli and Siniscalchi (1999), we show that there exists a continuous function $\pi_i : T_i \to \bigtimes_{h\in H_i} \Delta\big(\Theta_i \times S_j(h) \times T_j\big)$ such that for every $(\nu^h_i)_{h\in H_i} \in \bigtimes_{h\in H_i} \Delta\big(\Theta_i \times S_j(h) \times T_j\big)$ with $\nu^h_i\big(\{\theta_i\} \times S_j(h) \times T_{\theta_j})\big) = g^h_i(\theta_i)(\theta_j)$ for some $\theta_i \in \Theta_i$ and for all $h \in H_i$ and all $\theta_j \in \Theta_j$, there exists some $t_i \in T_{\theta_i}$ such that $\pi^h_i(t_i) = \nu^h_i$ for all $h \in H_i$. Finally, define the type structure $\mathcal{T}_{\mathfrak{F}} = \big((T_i)_{i\in I}, (\phi_i)_{i\in I}, (\lambda_i)_{i\in I}\big)$, by letting $\phi_i(t_{\theta_i}) := \theta_i$ and $\lambda^h_i(t_{\theta_i}) := \mathrm{marg}_{S_j(h)\times T_j} \pi^h_i(t_i)$ for every $\theta_i \in \Theta_i$ and every $h \in H_i$. Obviously, $\mathcal{T}_{\mathfrak{F}}$ is complete.

B. Proofs

B.1. Proofs of Section 4

We first introduce some additional notation and prove some intermediate results that we will use throughout the proof of our main theorem. Throughout the entire section, without loss of generality we consider a given structure $\mathfrak{F}$ such that for each $i \in I$, each $\theta_i \in \Theta_i$ and each $h \in H_i$ there exists a unique $\theta_{-i} \in \Theta_{-i}$ with $g^h_i(\theta_i)(\theta_{-i}) = 1$.

Lemma B1 (Optimality principle)**.** *Fix a structure $\mathfrak{F}$, an arbitrary player $i \in I$, an arbitrary $\theta_i \in \Theta_i$, an arbitrary history $h \in H_i$ and some $k > 0$. Then, a strategy $s_i \in S_i(h)$ is rational in $\big(B^k_{\theta_i}(h), S_i(h)\big)$ if and only if it is rational in $\big(B^k_{\theta_i}(h), D^{k-1}_{\theta_i}(h)\big)$.*

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

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

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

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

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

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

$$D_{\theta_i}^k(G) := \{s_i \in S_i : s_i \in D_{\theta_i}^k(h) \text{ for all } h \in H_i(s_i) \cap G\} \tag{B.3}$$

$$\begin{aligned} R_{\theta_i}^k(G) &:= \{s_i \in S_i : \text{there is } t_i \in T_{\theta_i}^k \text{ such that } (s_i, t_i) \in R_i^h \text{ for all } h \in H_i(s_i) \cap G\} \\ &= \mathrm{Proj}_{S_i}\Big(R_i^G \cap \big(S_i \times T_{\theta_i}^k\big)\Big). \end{aligned} \tag{B.4}$$

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

$$D_{\theta_i}(G) := \bigcap_{k=1}^{\infty} D_{\theta_i}^k(G).$$

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

$$R_{\theta_i}(G) := \bigcap_{k=1}^{\infty} R_{\theta_i}^k(G).$$

Construction of conditional beliefs. Fix an arbitrary $G \in \mathcal{H}$, an arbitrary $\theta_i \in \Theta_i$ and an arbitrary $s_i \in D_{\theta_i}^1(G)$. 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 $\beta_{\theta_i, s_i, G}^h \in \Delta\big(S_{-i}(h)\big)$ such that

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

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

- Suppose there exists some $k \in \mathbb{N}$ such that $s_i \in D_{\theta_i}^k(G) \setminus D_{\theta_i}^{k+1}(G)$. Then, it follows by definition that s_i is rational in $\big(B_{\theta_i}^k(h), D_{\theta_i}^{k-1}(h)\big)$. Hence, it follows from the optimality principle (Lemma B1) that we can choose some $\beta_{\theta_i, s_i, G}^h \in \Delta\big(B_{\theta_i}^k(h)\big)$ satisfying Eq. (B.5).

- Suppose that $s_i \in D^k_{\theta_i}(G)$ for all $k \in \mathbb{N}$. Then, it follows by definition that s_i is rational in $\big(B^k_{\theta_i}(h), D^{k-1}_{\theta_i}(h)\big)$ for every $k \in \mathbb{N}$. Thus we can choose some $\beta^h_{\theta_i,s_i,G} \in \Delta\big(B^{\mathfrak{F}}_{\theta_i}(h)\big)$ satisfying Eq. (B.5).

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

Construction of types. For each player $i \in I$ and each $\theta_i \in \Theta_i$ define the finite set

$$\Psi_{\theta_i} := \{\psi_{\theta_i,s_i,G} \mid (s_i, G) \in S_i \times \mathcal{H}\},$$

and let $\Psi_i := \bigcup_{\theta_i \in \Theta_i} \Psi_{\theta_i}$ and $\Psi_{-i} := \bigtimes_{j \neq i} \Psi_j$. Now, define the function $\phi_i : \Psi_i \to \Theta_i$ by $\phi_i(\psi_i) = \theta_i$ for each $\psi_i \in \Psi_{\theta_i}$. Moreover, define the mapping $\gamma^h_i : \Psi_i \to \Delta\big(S_{-i}(h) \times \Psi_{-i}\big)$ for each $h \in H_i$ as follows: For each $s_i \in D^1_{\theta_i}(G)$, let

$$\gamma^h_i(\psi_{\theta_i,s_i,G})(s_{-i}, \psi_{-i}) := \begin{cases} \beta^h_{\theta_i,s_i,G}(s_{-i}) & \text{if } \psi_j = \psi_{\theta_j,s_j,F_{\theta_i}(h)} \text{ for all } j \neq i \text{ and } g^h_i(\theta_i)\big((\theta_j)_{j \neq i}\big) = 1 \\ 0 & \text{otherwise.} \end{cases}$$

On the other hand, if $s_i \notin D^1_{\theta_i}(G)$, let $\gamma^h_i(\psi_{\theta_i,s_i,G})$ be an arbitrary probability measure over $S_{-i}(h) \times \Psi_{-i}$ such that $\big(\mathrm{marg}_{\Psi_{-i}} \gamma^h_i(\psi_{\theta_i,s_i,G})\big)\big(\bigtimes_{j \neq i} \phi^{-1}_j(\theta_j)\big) = g^h_i(\theta_i)\big((\theta_j)_{j \neq i}\big)$ for all $(\theta_j)_{j \neq i} \in \Theta_{-i}$. Now, observe that $\big((\Psi_i)_{i \in I}, (\phi_i)_{i \in I}, (\gamma_i)_{i \in I}\big)$ is a finite type structure, implying that each $\psi_i \in \Psi_i$ is associated with a hierarchy of conditional beliefs.

Recall that we have assumed $\mathcal{T}_{\mathfrak{F}} = \big((T_i)_{i \in I}, (\phi_i)_{i \in I}, (\lambda_i)_{i \in I}\big)$ to be a complete type structure. Then, it follows from Appendix A that there exists a function $\xi_i : \Psi_i \to T_i$ mapping each $\psi_{\theta_i,s_i,G}$ to a (unique) type $t_{\theta_i,s_i,G} := \xi_i(\psi_{\theta_i,s_i,G}) \in T_{\theta_i}$ such that (i) $t_{\theta_i,s_i,G}$ and $\psi_{\theta_i,s_i,G}$ induce the same conditional belief hierarchy, and moreover (ii) it is the case that $\phi_i(\psi_{\theta_i,s_i,G}) = \phi_i(t_{\theta_i,s_i,G})$. Furthermore, notice that by construction it is the case that $\lambda^h_i(t_{\theta_i,s_i,G})(s_{-i}, t_{-i}) = \gamma^h_i(\psi_{\theta_i,s_i,G})\big(s_{-i}, \xi^{-1}_i(t_{-i})\big)$ for all $(s_{-i}, t_{-i}) \in S_{-i} \times T_{-i}$. Finally, by construction it is the case that $(s_i, t_{\theta_i,s_i,G}) \in R^G_{\theta_i}$ whenever $s_i \in D^1_{\theta_i}(G)$.

Before moving on, for notation simplicity, let us adopt the convention that $T^0_{\theta_i}(h) := T_{\theta_i}$.

Lemma B2. *For every $i \in I$, every $\theta_i \in \Theta_i$, every $G \in \mathcal{H}$ and every $k > 0$, the following hold:*

(i) If $t_i \in T^{k-1}_{\theta_i}(h)$ then $b^h_i(t_i) \in \Delta\big(B^k_{\theta_i}(h)\big)$.

(ii) If $s_i \in D^k_{\theta_i}(G)$ then $t_{\theta_i,s_i,G} \in T^{k-1}_{\theta_i}(h)$ for all $h \in H_i(s_i) \cap G$.

(iii) $R^{k-1}_{\theta_i}(G) = D^k_{\theta_i}(G)$.

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 $B^{\mathfrak{F},1}_{\theta_i}(h) = B^{\mathfrak{F},0}_{\theta_i}(h) = S_{-i}(h)$, and therefore $\Delta\big(B^{\mathfrak{F},1}_{\theta_i}(h)\big) = \Delta\big(S_{-i}(h)\big)$, thus implying that $b^h_i(t_i) \in \Delta\big(B^{\mathfrak{F},1}_{\theta_i}(h)\big)$ for all $t_i \in T_{\theta_i}$, which proves (i). Moreover, recall from our convention that $T^{\mathfrak{F},0}_{\theta_i}(h) = T_{\theta_i}$, thus implying that $t_{\theta_i,s_i,G} \in T^{\mathfrak{F},0}_{\theta_i}(h)$ for all $h \in H_i(s_i) \cap G$, irrespective of whether $s_i \in D^{\mathfrak{F},1}_{\theta_i}$ or not, which proves (ii). Finally, notice that

$$\begin{aligned}
R^{\mathfrak{F},0}_{\theta_i}(G) &= \{s_i \in S_i : \text{there is } t_i \in T^{\mathfrak{F},0}_{\theta_i}(h) \text{ such that } (s_i, t_i) \in R^h_i \text{ for all } h \in H_i(s_i) \cap G\} \\
&= \{s_i \in S_i : \text{there is } t_i \in T_{\theta_i} \text{ such that } (s_i, t_i) \in R^G_i\} \\
&= D^{\mathfrak{F},1}_{\theta_i}(G)
\end{aligned}$$

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 $t_i \in T^k_{\theta_i}(h)$. Then, by definition it is the case that

$$t_i \in SB^h_{\theta_i}\big(R^{F_{\theta_i}(h)}_{-i} \cap (S_{-i} \times T^{k-1}_{-\theta_i}(F_{\theta_i}(h)))\big).$$

Then, we consider the following two cases:

(a) Let $R^{F_{\theta_i}(h)}_{-i} \cap (S_{-i} \times T^{k-1}_{-\theta_i}(F_{\theta_i}(h))) \neq \emptyset$.

By the definition of strong belief (at h) it is the case that $\lambda^h_i(t_i)\big(R^{F_{\theta_i}(h)}_{-i} \cap (S_{-i} \times T^{k-1}_{-\theta_i}(F_{\theta_i}(h)))\big) = 1$. Now, recall by Eq. (B.4) that

$$R^{k-1}_{-\theta_i}(F_{\theta_i}(h)) = \mathrm{Proj}_{S_{-i}}\big(R^{F_i(h)}_{-i} \cap (S_{-i} \times T^{k-1}_{-\theta_i}(F_{\theta_i}(h)))\big),$$

and therefore it follows that $b^h_i(t_i)\big(R^{k-1}_{-\theta_i}(F_{\theta_i}(h))\big) = 1$. Now observe that

$$\begin{aligned}
R^{k-1}_{-\theta_i}\big(F_{\theta_i}(h)\big) &= \bigtimes_{j \neq i} \big\{\, s_j \in S_j : s_j \in R^{k-1}_{\theta_j}\big(F_{\theta_i}(h)\big) \,\big\} \\
&= \bigtimes_{j \neq i} \big\{\, s_j \in S_j : s_j \in D^{k}_{\theta_j}\big(F_{\theta_i}(h)\big) \,\big\} && \big(\text{by the IA}\big) \\
&= \bigtimes_{j \neq i} \big\{\, s_j \in S_j : s_j \in D^{k}_{\theta_j}(h') \text{ for all } h' \in H_j \cap F_{\theta_i}(h) \,\big\}, && \text{(B.6)}
\end{aligned}$$

with $(\theta_j)_{j \neq i} \in \Theta_{-i}$ being such that $g^h_i(\theta_i)\big((\theta_j)_{j \neq i}\big) = 1$. Thus, it is the case that

$$\begin{aligned}
C^k_{\theta_i}(h) &= \bigtimes_{j \neq i} \big\{\, s_j \in S_j(h) : s_j \in D^{k}_{\theta_j}(h') \text{ for all } h' \in H_j \cap F_{\theta_i}(h) \,\big\} \\
&= S_{-i}(h) \cap R^{k-1}_{-\theta_i}\big(F_{\theta_i}(h)\big). && \text{(B.7)}
\end{aligned}$$

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

$$\begin{aligned} B^{k+1}_{\theta_i}(h) &= C^k_{\theta_i}(h) \\ &= S_{-i}(h) \cap R^{k-1}_{-\theta_i}\big(F_{\theta_i}(h)\big). \end{aligned}$$

Then, by combining $b^h_i(t_i)\big(R^{k-1}_{-\theta_i}(F_{\theta_i}(h))\big) = 1$ with $b^h_i(t_i)\big(S_{-i}(h)\big) = 1$, it is straightforward to obtain $b^h_i(t_i)\big(B^{k+1}_{\theta_i}(h)\big) = 1$. According to the second possibility we have $C^k_{\theta_i}(h) = \emptyset$, in which case we obtain $B^{k+1}_{\theta_i}(h) = B^k_{\theta_i}(h)$. But then, since $t_i \in T^k_{\theta_i}(h) \subseteq T^{k-1}_{\theta_i}(h)$, it follows from the IA that $b^h_i(t_i)\big(B^{k+1}_{\theta_i}(h)\big) = b^h_i(t_i)\big(B^k_{\theta_i}(h)\big) = 1$, which completes this part of the proof for the first case.

(b) Let $R^{F_{\theta_i}(h)}_{-i} \cap (S_{-i} \times T^{k-1}_{-\theta_i}(F_{\theta_i}(h))) = \emptyset$.

Then, it follows by definition that

$$\begin{aligned} R^{k-1}_{-\theta_i}\big(F_{\theta_i}(h)\big) \cap S_{-i}(h) &\subseteq R^{k-1}_{-\theta_i}\big(F_{\theta_i}(h)\big) \\ &= \mathrm{Proj}_{S_{-i}}\Big(R^{F_{\theta_i}(h)}_{-i} \cap (S_{-i} \times T^{k-1}_{-\theta_i}(F_{\theta_i}(h)))\Big) \\ &= \emptyset \end{aligned} \tag{B.8}$$

Now, using the same reasoning as in Eq. (B.6), combined with Eq. (B.8), we obtain

$$\begin{aligned} R^{k-1}_{-\theta_i}(F_{\theta_i}(h)) \cap S_{-i}(h) &= \bigtimes_{j \neq i} \big\{ \, s_j \in S_j(h) : s_j \in D^k_{\theta_j}(h') \text{ for all } h' \in H_j \cap F_{\theta_i}(h) \, \big\} \\ &= \emptyset, \end{aligned}$$

again with $(\theta_j)_{j\neq i} \in \Theta_{-i}$ being such that $g^h_i(\theta_i)\big((\theta_j)_{j\neq i}\big) = 1$. Moreover, using the same argument as in Eq. (B.7), we obtain

$$C^k_{\theta_i}(h) = S_{-i}(h) \cap R^{k-1}_{-\theta_i}\big(F_{\theta_i}(h)\big).$$

Thus, combining the previous two equations, we conclude that $C^k_{\theta_i}(h) = \emptyset$. Hence, $B^{k+1}_{\theta_i}(h) = B^k_{\theta_i}(h)$. Finally, since $t_i \in T^k_{\theta_i}(h) \subseteq T^{k-1}_{\theta_i}(h)$, it follows from the IA that $b^h_i(t_i)\big(B^{k+1}_{\theta_i}(h)\big) = b^h_i(t_i)\big(B^k_{\theta_i}(h)\big) = 1$, which completes the proof of part (i).

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

$$t_{\theta_i,s_i,G} \in SB^h_{\theta_i}\Big(R^{F_{\theta_i}(h)}_{-i} \cap \big(S_{-i} \times T^{k-1}_{-\theta_i}(F_{\theta_i}(h))\big)\Big). \tag{B.9}$$

The latter amounts to proving that

$$\Big[\; R^{F_{\theta_i}(h)}_{-i} \cap \big(S_{-i} \times T^{k-1}_{-\theta_i}(F_{\theta_i}(h))\big) \neq \emptyset \;\Big] \Rightarrow \Big[\; \lambda^h_i(t_{\theta_i,s_i,G})\big(R^{F_{\theta_i}(h)}_{-i} \cap \big(S_{-i} \times T^{k-1}_{-\theta_i}(F_{\theta_i}(h))\big)\big) = 1 \;\Big] \tag{B.10}$$

First, notice that $t_{\theta_i,s_i,G} \in SB^h_{\theta_i}\big(R^{F_{\theta_i}(h)}_{-i} \cap \big(S_{-i} \times T^{k-1}_{-\theta_i}(F_{\theta_i}(h))\big)\big)$ is trivially satisfied whenever $R^{F_{\theta_i}(h)}_{-i} \cap \big(S_{-i} \times T^{k-1}_{-\theta_i}(F_{\theta_i}(h))\big) = \emptyset$. Hence, we will focus on the case where $R^{F_{\theta_i}(h)}_{-i} \cap \big(S_{-i} \times T^{k-1}_{-\theta_i}(F_{\theta_i}(h))\big) \neq \emptyset$. Recall that $\big((\theta_j)_{j\neq i}\big)$ is the unique element of Θ_{-i} receiving positive probability by $g^h_i(\theta_i)$. Then, for every $j \neq i$, there exists some $(s^*_j, t^*_j) \in S_j(h) \times T_{\theta_j}$ such that (i) $(s^*_j, t^*_j) \in R^{h'}_j$ for all $h' \in H_j(s^*_j) \cap F_{\theta_i}(h)$, and (ii) $t^*_j \in T^{k-1}_{\theta_j}(h')$ for all $h' \in H_j \cap F_{\theta_i}(h)$.

Now, we are going to prove that $s^*_j \in D^k_{\theta_j}(h')$ for every $h' \in H_j(s^*_j) \cap F_{\theta_i}(h)$. To do so, take an arbitrary $t^{k-1}_j \in T^{k-1}_{\theta_j}$, and define the type t^{**}_j by

$$\lambda^{h'}_j(t^{**}_j) := \begin{cases} \lambda^{h'}_j(t^*_j) & \text{for each } h' \in H_j(s^*_j) \cap F_{\theta_i}(h), \\ \lambda^{h'}_j(t^{k-1}_j) & \text{for each } h' \in H_j \setminus \big(H_j(s^*_j) \cap F_{\theta_i}(h)\big). \end{cases}$$

Notice that since $\mathcal{T}_{\mathfrak{F}}$ is a complete type structure, such a type exists. Observe that by construction it is the case that $(s^*_j, t^{**}_j) \in R^{F_{\theta_i}(h)}_j$, and moreover $t^{**}_j \in T^{k-1}_{\theta_j}$. Therefore, we obtain

$$\begin{aligned} s^*_j &\in R^{k-1}_{\theta_j}\big(F_{\theta_i}(h)\big) \cap S_j(h) \\ &= D^k_{\theta_j}\big(F_{\theta_i}(h)\big) \cap S_j(h) && \big(\text{by the IA}\big) \\ &= \big\{\ s_j \in S_j(h) : s_j \in D^k_{\theta_j}(h') \text{ for all } h' \in H_j(s_j) \cap F_{\theta_i}(h)\ \big\} \\ &\neq \emptyset. \end{aligned}$$

The latter implies directly by definition that $C^k_{\theta_i}(h) \neq \emptyset$. Hence, it is – also by definition – the case that

$$B^{k+1}_{\theta_i}(h) = C^k_{\theta_i}(h). \tag{B.11}$$

Now, notice that by construction $\lambda^h_i(t_{\theta_i,s_i,G})$ put positive probability only to strategy-type pairs (s_j, t_j) such that $t_j = t_{\theta_j,s_j,F_{\theta_i}(h)}$. Moreover, since $s_i \in D^{k+1}_{\theta_i}(G)$ it follows from the construction of the beliefs that $b^h_i(t_{\theta_i,s_i,G}) \in \Delta\big(B^{k+1}_{\theta_i}(h)\big)$. Therefore, it follows from Eq. (B.11) that $\text{marg}_{S_j \times T_j} \lambda^h_i(t_{\theta_i,s_i,G})$ puts positive probability only to strategy-type pairs $(s_j, t_j) \in S_j(h) \times T_j$ such that $t_j = t_{\theta_j,s_j,F_i(h)}$ and $s_j \in D^k_{\theta_j}(h')$ for all $h' \in H_j(s_j) \cap F_{\theta_i}(h)$. Hence, from the IA it follows that $\text{marg}_{S_j \times T_j} \lambda^h_i(t_{\theta_i,s_i,G})$ assigns probability 1 to

$$R^{F_{\theta_i}(h)}_j \cap \big\{\ (s_j, t_j) \in S_j \times T_j : t_j \in T^{k-1}_{\theta_j}(h') \text{ for all } h' \in H_j \cap F_{\theta_i}(h)\ \big\}$$

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

Proof of (iii): First, we prove that $R^{k-1}_{\theta_i}(G) \subseteq D^k_{\theta_i}(G)$: Take an arbitrary $s_i \in R^{k-1}_{\theta_i}(G)$. By definition there exists a type in $t_i \in T^{k-1}_{\theta_i}$ such that $(s_i, t_i) \in R^G_i$. Now, by part (i) of the result – that we have already proven above – it follows that $b^h_i(t_i)\big(B^k_{\theta_i}(h)\big) = 1$ 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

$\big(B^k_{\theta_i}(h), D^{k-1}_{\theta_i}(h)\big)$. Thus, we conclude that $s_i \in D^k_{\theta_i}(h)$ for all $h \in H_i(s_i) \cap G$. The latter directly implies that $s_i \in D^k_{\theta_i}(G)$ which completes this part of the proof.

Second, we prove that $D^k_{\theta_i}(G) \subseteq R^{k-1}_{\theta_i}(G)$: Take an arbitrary $s_i \in D^k_{\theta_i}(G)$. Then, by part (ii) that we have already proven above, it follows that $t_{\theta_i,s_i,G} \in T^{k-1}_{\theta_i}(h)$ for all $h \in G \cap H_i(s_i)$. Now, fix an arbitrary type $t^{k-1}_i \in T^{k-1}_{\theta_i}$, and define the type $t^*_{\theta_i,s_i,G} \in T_{\theta_i}$ by

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

Notice that since $\mathcal{T}_{\mathfrak{F}}$ is a complete type structure, such a type exists. Then, by construction it is the case that $t^*_{\theta_i,s_i,G} \in T^{k-1}_{\theta_i}$, and therefore it follows that $(s_i, t^*_{\theta_i,s_i,G}) \in R^h_i$ for all $h \in G \cap H_i(s_i)$. Hence, we conclude that $s_i \in R^{k-1}_{\theta_i}(G)$, which completes the proof of the lemma. $\square$

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

Proof of (i): It follows directly from Lemma B2.i.

Proof of (ii): Fix an arbitrary $\beta^h_i \in \Delta\big(B^{\mathfrak{F},k}_{\theta_i}(h)\big)$, and let $s^*_i \in D^{\mathfrak{F},k}_{\theta_i}(h)$ be such that

$$U^h_i(s^*_i, \beta^h_i) \geq U^h_i(s_i, \beta^h_i) \tag{B.12}$$

for all $s_i \in D^{\mathfrak{F},k-1}_{\theta_i}(h)$. In fact, notice that Eq. (B.12) holds, not only for every $s_i \in D^{\mathfrak{F},k-1}_{\theta_i}(h)$, but for every $s_i \in S_i(h)$ (see Lemma B1). Now, we define $\beta^h_{\theta_i,s^*_i,\{h\}} := \beta^h_i$, and construct the type $t_{\theta_i,s^*_i,\{h\}}$ like we did above. Then, by Lemma B2.ii, it is the case that $t_{\theta_i,s^*_i,\{h\}} \in T^{\mathfrak{F},k-1}_{\theta_i}(h)$, which – together with the fact that $\beta^h_{\theta_i,s^*_i,\{h\}} := b^h_i(t_{\theta_i,s^*_i,\{h\}})$ – completes the proof. $\square$

Proof of Theorem 2. Observe that by construction

$$\begin{aligned} R^{\mathfrak{F}}_{\theta_i}(H) &= \mathrm{Proj}_{S_i}\big(R_i \cap (S_i \times T^{\mathfrak{F}}_{\theta_i})\big) \\ D^{\mathfrak{F}}_{\theta_i}(H) &= \{s_i \in S_i : s_i \in D^{\mathfrak{F}}_{\theta_i}(h) \text{ for all } h \in H_i(s_i)\}, \end{aligned}$$

and recall by Lemma B2.iii that $R^{\mathfrak{F}}_{\theta_i}(H) = D^{\mathfrak{F}}_{\theta_i}(H)$, which completes the proof. $\square$

B.2. Proofs of Section 5

In this section, we focus on structures $\mathfrak{F}$ with commonly known $F \in \mathcal{F}$, implying that Θ_i is a singleton for each $i \in I$. Thus, recall that we identify the unique θ_i with i, e.g., we write $F_i(h)$ for $F_{\theta_i}(h)$.

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

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

Hence, it is the case that

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

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 $Q_{-i}^k(h) = \bigtimes_{j \neq i} Q_j^k(h)$. 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

$$Q_i(h) := \bigcap_{k=1}^{\infty} Q_i^k(h), \tag{B.13}$$

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 B3. *Let the structure $\mathfrak{F}$ be such that $(F_i)_{i \in I}$ is commonly known with $F_i(h) = \mathrm{Fut}(h)$ for all $i \in I$ and all $h \in H_i$. Then, for every $i \in I$, every $h \in H_i$ and every $k > 1$ the following hold:*

(i) $Q^k_{-i}(h) = B^k_i(h)$.

(ii) $Q^{k+1}_i(h) = \{s_i \in S_i(h) : s_i \in D^k_i(h') \text{ for all } h' \in \mathrm{Fut}(h) \cap H_i(s_i)\}$.

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

$$\begin{aligned}
B^k_i(h) &= C^{k-1}_i(h) && \big(\text{because } C^{k-1}_i(h) \neq \emptyset\big)\\
&= \bigtimes_{j \neq i} \{s_j \in S_j(h) : s_j \in D^{k-1}_j(h') \text{ for all } h' \in H_j(s_j) \cap \mathrm{Fut}(h)\}\\
&= \bigtimes_{j \neq i} Q^k_j(h) && \big(\text{by the IA}\big)\\
&= Q^k_{-i}(h),
\end{aligned}$$

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

Now, we move to the inductive step for part (ii). Again, fix an arbitrary $i \in I$ and an arbitrary $h \in H_i$, and take an arbitrary $s_i \in Q^{k+1}_i(h)$. Then, by definition, s_i is rational in $\big(Q^k_{-i}(h'), Q^k_i(h')\big)$ 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 $\big(B^k_i(h'), Q^k_i(h')\big)$ for every $h' \in \mathrm{Fut}(h) \cap H_i(s_i)$. Now, notice that for every $s'_i \in S_i(h')$,

$$\begin{aligned}
s'_i \text{ is rational in } \big(B^k_i(h'), Q^k_i(h')\big) \quad &\Leftrightarrow \quad s'_i \text{ is rational in } \big(B^k_i(h'), S_i(h')\big)\\
&\Leftrightarrow \quad s'_i \text{ is rational in } \big(B^k_i(h'), D^{k-1}_i(h')\big).
\end{aligned}$$

The first equivalence follows from Perea (2012, Lem. 8.14.6), while the second one follows from Lemma B1 above. Hence, s_i is rational in $\big(B^k_i(h'), D^{k-1}_i(h')\big)$ for every $h' \in \mathrm{Fut}(h) \cap H_i(s_i)$, thus implying that $s_i \in D^k_i(h')$ for every $h' \in \mathrm{Fut}(h) \cap H_i(s_i)$. Therefore,

$$Q^{k+1}_i(h) \subseteq \{s_i \in S_i(h) : s_i \in D^k_i(h') \text{ for all } h' \in \mathrm{Fut}(h) \cap H_i(s_i)\}. \tag{B.14}$$

Now, in order to prove the inverse weak inequality, take some $s_i \in D^k_i(h')$ for every $h' \in \mathrm{Fut}(h) \cap H_i(s_i)$. This implies that s_i is rational in $\big(Q^k_{-i}(h'), D^{k-1}_i(h')\big)$ for every $h' \in \mathrm{Fut}(h) \cap H_i(s_i)$, and by the previous sequence of equivalences, s_i is rational in $\left(Q^k_{-i}(h'), Q^k_i(h')\right)$ 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

$$Q_i^{k+1}(h) \supseteq \{s_i \in S_i(h) : s_i \in D_i^k(h') \text{ for all } h' \in \mathrm{Fut}(h) \cap H_i(s_i)\}. \tag{B.15}$$

Then, inequalities (B.14) and (B.15) complete this part of the proof. $\square$

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 $s_i \in \mathrm{Proj}_{S_i}\left(R_i \cap (S_i \times CFB_i)\right)$. Moreover, from our Theorem 2, a strategy s_i can be rationally played under $\mathfrak{F}$-CSBR (in a complete type structure) if and only if it survives the $\mathfrak{F}$-ICDP, i.e., formally, $s_i \in D_i^{\mathfrak{F}}(h)$ for all $h \in H_i(s_i)$ if and only if $s_i \in \mathrm{Proj}_{S_i}\left(R_i \cap (S_i \times T_i^{\mathfrak{F}})\right)$. Thus, it suffice to prove that a strategy survives BDP if and only if it survives $\mathfrak{F}$-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 B3, the latter is true if and only if $s_i \in \{s_i' \in S_i(h) : s_i' \in D_i^k(h') \text{ for all } h' \in \mathrm{Fut}(h) \cap H_i(s_i)\}$ for all $k > 0$ and for all $h \in H_i(s_i)$. Obviously, the latter is equivalent to $s_i \in D_i^k(h)$ for every $k > 0$ and every $h \in H_i(s_i)$, which by definition means that s_i survives the $\mathfrak{F}$-ICDP, thus completing the proof. $\square$

References

Aliprantis, C. & Border, K. (1994). *Infinite dimensional analysis*. Springer Verlag, Berlin.

Aumann, R.J. & Sorin, S. (1989). Cooperation and bounded recall. *Games and Economic Behavior* 1, 5–39.

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

Barlo, M., Carmona, M. & Sabourian, H. (2009). Repeated games with one-memory. *Journal of Economic Theory* 312–336.

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.

Bonanno, G. (2004). Memory and perfect recall in extensive games. *Games and Economic Behavior* 47, 237–256.

Brandenburger, A. & Dekel, E. (1993). Hierarchies of beliefs and common knowledge. *Journal of Economic Theory* 59, 189–198.

Chen, J. & Micali, S. (2013). The order independence of iterated dominance in extensive games. *Theoretical Economics* 8, 125–163.

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.

Cole, J. (2001). *Forget colonialism?* University of California Press.

Cole, H.L. & Kocherlakota, N.R. (2005). Finite memory and imperfect monitoring. *Games and Economic Behavior* 53, 59–72.

De Oliveira, H., Denti, T., Mihm, M. & Ozbek, K. (2014). Rationally inattentive preferences. *Working Paper.*

Dow, J. (1991). Search decisions with limited memory. *Review of Economic Studies* 58, 1–14.

Ellis, A. (2015). Foundations for optimal inattention. *Working Paper.*

Feinberg, Y. (2012). Games with unawareness. *Stanford Graduate School of Business Paper*, 2122.

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.

Grossi, D. & Turrini, P. (2012). Short sight in extensive games. *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, 805–812.

Halpern, J.Y. & Rêgo, L.C. (2014). Extensive games with possibly unaware players. *Mathematical Social Sciences* 70, 42–58.

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., Meier, M. & Schipper, B. (2013). Dynamic unawareness and rationalizable behavior. *Games and Economic Behavior* 81, 50–68.

Heifetz, A. & Perea, A. (2015). On the outcome equivalence of backward induction and extensive form rationalizability. *International Journal of Game Theory* 44, 37–59.

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

Jehiel, P. (2001). Limited foresight may force cooperation. *Review of Economic Studies* 68, 369–391.

Johnson, E.J., Camerer, C., Sen, S. & Rymon, T. (2002). Detecting failures of backward induction: monitoring information search in sequential bargaining. *Journal of Economic Theory* 104, 16–47.

Kline, J.J. (2002). Minimum memory for equivalence between ex ante optimality and time-consistency. *Games and Economic Behavior* 38, 278–305.

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

Lau, R.R. & Redlawsk, D.P. (2006). *How voters decide: Information processing in an election campaign*. Cambridge University Press.

Lehrer (1988). Repeated games with stationary bounded recall strategies. *Journal of Economic Theory* 46, 130–144.

Maenner, E. (2008). Adaptation and complexity in repeated games. *Games and Economic Behavior* 63, 166–187.

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

Mengel (2014). Learning by (limited) forward looking players. *Journal of Economic Behavior and Organization* 108, 59–77.

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

Penta, A. (2015). Robust dynamic implementation. *Journal of Economic Theory* (forthcoming).

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.

Piccione, M. & Rubinstein, A. (1997). On the interpretation of decision problems with imperfect recall. *Games and Economic Behavior* 20, 3–24.

Reny, P.J. (1992). Backward induction, normal form perfection and explicable equilibria. *Econometrica* 60, 627–649.

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.

——— (1998). *Modeling bounded rationality.* MIT Press.

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.

Sims, C. (2003). Implications of rational inattention. *Journal of Monetary Economics* 50, 665–690.

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

Turrini, P. (2015). Computing rational decisions in extensive games with limited foresight. *Proceedings of the 13th AAAI conference on Artificial Intelligence (AAAI 2016)* (forthcoming).