Rationalizability and Observability*

Antonio Penta† Peio Zuazo-Garin‡

March 20, 2018

**Abstract**

We study the strategic impact of players' higher order uncertainty over whether their actions are observable to their opponents. We characterize the predictions of Rationality and Common Belief in Rationality (RCBR) which are 'robust' in the sense that they do not depend on the restrictions on players' infinite order beliefs over the extensive form (that is, the observability of actions). We show that RCBR is generically unique, and that its robust predictions often support a robust refinement of rationalizability. For instance, in unanimity games, the robust predictions of RCBR rule out any inefficient equilibrium action; in zero-sum games, they support the maxmin solution, resolving a classical tension between RCBR and the maxmin logic; in common interest games, RCBR generically ensures efficient coordination of behavior, thereby showing that higher order uncertainty over the extensive form serves as a mechanism for equilibrium coordination on purely eductive grounds.

We also characterize the robust predictions in settings with asynchronous moves, but in which the second mover does not necessarily observe the first mover's action. In these settings, higher order uncertainty over the observability of the earlier choice yields particularly sharp results: in 'Nash-commitment games', for instance, RCBR generically selects the equilibrium of the static game which is most favorable to the earlier mover. This means that a first-mover advantage arises whenever higher-order beliefs do not rule out that it might exist, even if the earlier mover's action is not observable. Hence, in the presence of extensive form uncertainty, timing alone may detemine the attribution of the strategic advantage, independent of the actual observability of choices.

**Keywords**: coordination – extensive form – first-mover advantage – higher order beliefs – rationalizability – robustness

# 1 Introduction

A large literature in game theory has studied the effects of perturbing common knowledge assumptions on payoffs, from different perspectives (e.g., Rubinstein (1989), Carlsson and

*We thank Pierpaolo Battigalli, Eddie Dekel, Laura Doval, Drew Fudenberg, Lones Smith, Stephen Morris, Bill Sandholm, Joel Sobel, Jakub Steiner, Joel Watson and Muhamet Yildiz for helpful comments and suggestions. We also thank seminar audiences at the Univ. Pública de Navarra, Cambridge, UW-Madison, Univ. of the Basque Country, PSE, CERGE-EI , UPF, as well as the participants of the 2017 SISL Summer Conference (Caltech), the 2017 HKU-UBC Summer Workshop in Economic Theory (Hong-Kong Univ.), the StratEmon Workshop (Bocconi), and the 2018 Southampton-Bristol Economic Theory Workshop (Southampton). Antonio Penta acknowledges the financial support of the ERC Starting Grant #759424.

†UW-Madison, Universitat Pompeu Fabra and Barcelona-GSE. E-mail: antonio.penta@upf.edu

‡University of the Basque Country (UPV/EHU). E-mail: peio.zuazo@ehu.eus

van Damme (1993), Kaji and Morris (1997), Morris and Shin (1998), Weinstein and Yildiz (2007, 2011, 2012, 2016), etc.). In contrast, the assumption of common knowledge of the extensive form has hardly been challenged.[1] Yet, uncertainty over the extensive form is key to many strategic situations. It is clearly paramount in military applications, but economic settings abound in which players are uncertain over the moves that are available to their opponents, or over their information about earlier moves, etc.

For instance, when we study firms interacting in a market, we often model the situation as a static game (Cournot, Bertrand, etc.), or as a dynamic one (e.g., Stackelberg). But, in the former case, this not only presumes that firms' decisions are made without observing the choices of their competitors, but also that this is common knowledge among them. Yet, firms in reality may often be concerned that their decisions could be leaked to their competitors. Or perhaps consider that other firms may be worried about that, or that their competitors may think the same about them, and so on. In other words, firms may face higher order uncertainty over the extensive form (whether the game truly is static, or whether actions are leaked) in ways which would be impossible to model with absolute precision.[2] It is then natural to ask which predictions we can make, using standard models (and hence abstracting from the fine details of such belief hierarchies), which would remain valid even if players' beliefs over the extensive form were misspecified in our model.

The problem of extensive-form robustness is broad and challenging, and could be approached from many angles. In this paper we study the strategic impact of players' higher order uncertainty over the observability of their actions.[3] More precisely, we focus on two-player games, and consider the space of all belief hierarchies generated by a fundamental space of uncertainty in which players are unsure whether the game will be played as a static game, i.e. with no information about each other's move, or sequentially, with perfect information. We then characterize the strongest predictions that are robust to perturbing players' higher order beliefs over the extensive form. Our main focus is to understand how game theoretic predictions for a static game are affected when players do choose simultaneously, but they may face higher order uncertainty over the possibility that their actions would be leaked to their opponent. But our results also cover general situations in which moves may be simultaneous or sequential and players are uncertain about it at any order of beliefs. We show that the 'strongest robust predictions' coincide with those of Rationality and Common Belief in Rationality (RCBR) generically on the universal type space, that RCBR is generically unique and that its robust predictions support a proper refinement of (static) rationalizability.

The generic uniqueness result is reminiscent of the famous result from Weinstein and

[1] As we discuss in Sections 1.2 and 8, some papers have studied commonly known structures to represent players' uncertainty over the extensive form, but none of these papers has relaxed common knowledge assumptions in the sense that we do here, or in the works on payoff uncertainty mentioned above.

[2] The discussion applies just as well to dynamic models (e.g., a Stackelberg duopoly): even if the leader's action is observable to the follower, and hence we may decide to model the situation as a game with perfect information, in practice it may be that the leader may worry that the follower has already secretely committed to an action, or that he may think the leader is uncertain about that, etc. In other words, even if the leader's action is observable, it need not be common knowledge.

[3] The broad question of extensive form robustness is gaining increasing attention in the literature (e.g., Ely and Doval (2016), Salcedo (2017), Makris and Renou (2018) and, in mechanism design, Peters (2015). These works, which consider related but distinct notions of robustness, are discussed in Sections 1.2 and 8.

Yildiz (2007), who considered perturbations of common knowledge assumptions on payoffs. The similarity, however, is only superficial, as the impact of higher order uncertainty over the observability of actions is very different from that of payoff uncertainty. In particular, Weinstein and Yildiz show that whenever a type admits multiple Interim Correlated Rationalizable (ICR, Dekel, Fudenberg and Morris (2007)) actions, for any of such actions there exists another type with arbitrarily close belief hierarchy for which that action is uniquely rationalizable. It follows that no refinement of ICR is robust, when higher order uncertainty on payoffs is introduced. In contrast, our results generally support a robust refinement of rationalizability. For instance, in 'unanimity games', the robust predictions rule out any inefficient equilibrium action; in zero-sum games, they uniquely select the maxmin solution, solving a tension between RCBR and the maxmin logic which has long been discussed in the literature (e.g., von Neumann and Morgenstern (1947, Ch.17), Luce and Raiffa (1957, Ch.4), Schelling (1960, Ch.7), etc.); in *Nash commitment* games with a unique efficient equilibrium (which include as special cases both zero-sum games with a pure equilibrium and Aumann and Sorin's (1989) common interest games), RCBR generically ensures efficient coordination of behavior, thereby showing that higher order uncertainty over the extensive form may also serve as a mechanism for equilibrium coordination based on purely 'eductive' grounds (Binmore (1987-88); see also Guesnerie (2005)).[4] Thus, the overall message is radically different from Weinstein and Yildiz's unrefinability result.

We also characterize the robust predictions under 'asymmetric' perturbations, in the sense that we maintain common knowledge that one player's action is *not* observable, but there may be higher order uncertainty over the observability of the other player's action. Asymmetric uncertainty arises naturally in a number of settings, for instance when moves are chosen at different points in time, with a commonly known order. Asymmetric perturbations therefore can be used (*i*) to study the robustness of game theoretic predictions for perfect information games, when players may entertain higher order uncertainty over whether the first mover's action will indeed be observed; but also (*ii*) to study the robustness of game theoretic predictions for static games with asynchronous moves, in which players may face higher order uncertainty over whether the earlier choice is leaked to the second mover.

In these settings, the analysis delivers particularly striking results. In Nash commitment games, for instance, we show that RCBR generically selects the equilibrium of the static game which is most favorable to the earlier mover, regardless of whether his action is actually observable. Hence, a first-mover advantage arises in these games whenever higher order beliefs do not rule out its existence. Such pervasiveness of the first-mover advantage with asynchronous moves has obvious relevance from a strategic viewpoint. It suggests that, in the presence of higher order uncertainty on the observability of actions, *timing* of moves alone may determine the attribution of the strategic advantage.[5] This message is clearly at odds

[4]The term 'eductive' was introduced by Binmore (1987-88), to refer to the rationalistic, reasoning-based approach to the foundations of solution concepts. It was contrasted with the 'evolutive approach', in which solution concepts are interpreted as the steady state of an underlying learning or evolutive process. The question of eductive stability has been pursued in economics both in partial and general equilibrium settings (see, e.g., Guesnerie (2005) and references therein).

[5]Asynchronous moves in coordination games have also been studied by Lagunoff and Matsui (1997) and Ambrus and Ishii (2015), in a repeated game setting in which choices are reversible but also observable. (On

with the received game theoretic intuition that commitment and observability, not timing, are key to ensure the upper hand in a strategic situation. Our results show that this classical insight is somewhat fragile, and in fact overturned, when one considers even arbitrarily small departures from the standard assumptions of common knowledge on the extensive form.

A large experimental literature has explored the impact of timing on individuals' choices in a static game, with findings that are often difficult to reconcile with the received game theoretic wisdom. For instance, asynchronous moves in the Battle of the Sexes systematically select the Nash equilibrium most favorable to the first mover (e.g., Camerer (2003), Ch.7), thereby confirming an earlier conjecture by Kreps (1990), who also pointed at the difficulty of making sense of this intuitive idea in a classical game theoretic sense:

> "From the perspective of game theory, the fact that player B moves first chronologically is not supposed to matter. It has no effect on the strategies available to players nor to their payoffs. [...] however, and my own casual experiences playing this game with students at Stanford University suggest that in a surprising proportion of the time (over 70 percent), players seem to understand that the player who 'moves' first obtains his or her preferred equilibrium. [... ] And *formal mathematical game theory has said little or nothing about where these expectations come from, how and why they persist, or when and why we might expect them to arise.*" (Kreps, 1990, pp.100-101 (italics in the original)).

Our results achieve this goal, as they show that the behavior observed in these experiments is precisely the unique robust prediction of RCBR, when one considers higher order uncertainty over the observability of actions. This is not to say that the logic of our results – which involves possibly very complex higher order reasoning – necessarily provides a behaviorally accurate model of strategic thinking (see, for instance, Crawford, Costa-Gomes and Iriberri (2013)). Nonetheless, our results imply that, once appended with this kind of uncertainty, standard assumptions such as RCBR may provide an effective *as if* model of how timing impacts players' strategic reasoning. This has clear methodological implications for the behavioral and experimental literature, which we further discuss in Section 7.

The rest of the paper is organized as follows: Section 1.1 presents a leading example, and Section 1.2 reviews the most closely related literature; Section 2 introduces the model; Section 3 formalizes the notion of RCBR under extensive-form uncertainty; Section 4 contains our main result, Theorem 1, which provides a structure theorem for RCBR and characterizes its robust predictions. We then explore some of Theorem 1's implications for robust refinements (Section 5) and for eductive coordination (Section 6). Section 7 provides the analysis of the asymmetric perturbations. Section 8 concludes.

---

reversibility of actions, see also Kovac and Steiner (2013)). In our setting, coordination arises from higher order uncertainty over the extensive form, independent of the actual observability of the earlier mover's action. In this sense, our exercise is purely about asynchronicity, independent of the actual observability.

## 1.1 Leading Example

We begin with a simple example to illustrate the basic elements of our model and some of the main differences from Weinstein and Yildiz's (2007). Consider the following game (an augmented Battle of the Sexes, with one additional inefficient equilibrium):

| | $L$ | $C$ | $R$ |
|---|---|---|---|
| $U$ | 4 2 | 0 0 | 0 0 |
| $M$ | 0 0 | 2 4 | 0 0 |
| $D$ | 0 0 | 0 0 | 1 1 |

The (pure) Nash equilibria are on the main diagonal. The equilibrium $(D, R)$ is inefficient, whereas $(U, L)$ and $(M, C)$ are both efficient, but the two players (*Ann*, or $A$, the row player, and *Bob*, or $B$, the column player) have conflicting preferences over which equilibrium they would like to coordinate on. Clearly, everything is rationalizable in this game. Hence, as shown by Weinstein and Yildiz (2007), if – maintaining common knowledge that the game is static – we relax all common knowledge assumptions on payoffs and perturb higher order beliefs about payoffs, then any action is uniquely rationalizable (and hence the unique equilibrium action) for some types that are arbitrarily close to having common belief that payoffs are as in the game above. It follows that the only predictions that are robust to possible mis-pecifications of higher order beliefs on payoffs are those provided by rationalizability, which does not rule anything out in this game.

Here we consider a different kind of perturbations of the common knowledge assumptions. That is, we maintain that payoffs are common knowledge, but we introduce higher order uncertainty over the observability of players' actions. In particular, suppose that there are three states of the world, $\omega^0$, $\omega^A$ and $\omega^B$: in $\omega^0$, the game is static; in $\omega^i$ instead, for $i = A, B$, player $i$ moves first, and then $j \neq i$ chooses having observed $i$'s move. Clearly, if the true state is $\omega^A$, and this is commonly knowledge, the only strategy profile consistent with players' (sequential) Rationality and Common Belief in Rationality (RCBR) is the backward induction solution, which induces outcome $(U, L)$, the equilibrium of the static game most favorable to player 1. Similarly, if $\omega^B$ were common knowledge, the only outcome consistent with RCBR would be $(C, M)$.

Now imagine a situation in which the game is actually static (i.e., the true state is $\omega^0$), and both players know that, but Bob thinks that Ann thinks it common belief that the state is $\omega^A$. Then, he expects her to choose $U$, and hence choosing $L$ is his only best reply. Moreover, if Ann believes that Bob's beliefs are just as described, she also picks $U$ as the only action consistent with RCBR. But then, if Bob believes that Ann so believes, his unique best response is to indeed play $L$, and so on. Iterating this argument, it is easy to see that we can have Ann and Bob share arbitrarily many levels of mutual belief that the game is static, and yet have $(U, L)$ as the only outcome consistent with RCBR. No matter how close we get to common belief of $\omega^0$, as long as higher order beliefs do not rule out the possibility of $\omega^A$, the only outcome consistent with RCBR is $(U, L)$, which is *as if* Ann and Bob commonly believed

$\omega^A$, not $\omega^0$. Or, more evocatively: *Ann* de facto *has a first-mover advantage, whenever she is believed to have it at some arbitrarily high order of beliefs.*

Clearly, if higher order beliefs traced back to $\omega^B$ instead of $\omega^A$, the logic above would uniquely select $(M, C)$, the equilibrium most favorable to Bob. But it can be shown that no such perturbation would robustly select actions $D$ and $R$. Hence, we obtain at least two insights from this example, which we will show to hold in a large class of games: $(i)$ first, higher order uncertainty over the observability of actions rules out the inefficient equilibrium actions (the general result is in fact stronger than this); $(ii)$ second, if we maintain common knowledge that Ann *does not* observe Bob's move (for instance, because it is common knowledge that she moves earlier), then higher order uncertainty over the observaility of actions gives Ann a full first-mover advantage, regardless of the actual observability of her action.

Results $(i)$-$(ii)$ in this example are driven by the fact that, given the nature of extensive-form uncertainty, only the two backward induction outcomes could be used to start the 'infection argument' (or just $(U, L)$, in the case of $(ii)$). But this is only one of the differences between our analysis and Weinstein and Yildiz's. Because of the particular configuration of payoffs, the infection argument in the example above only involved a standard chain of (static) best responses. In general games, however, the robust predictions also depend on the behavior of types who are uncertain over whether the game is static or dynamic, whose optimization problem therefore is a hybrid of the standard static and dynamic ones. These hybrid best responses carry over to the higher order beliefs, and hence also the way the infection spreads from one type to another will differ from Weinstein and Yildiz's. As a result, the characterization of the robust predictions in our setting will be quite different from theirs. These further diferences will be explained below.

## 1.2 Related Literature

The closest papers to our work are those which study perturbations of common knowledge assumptions on payoffs, following the seminal paper by Weinstein and Yildiz (2007). Weinstein and Yildiz (2007) characterize the correspondence of ICR on the univeral type space generated by a space of payoff uncertainty which satisfies a richness condition for static games. The key insights have been applied to mechanism design by Oury and Tercieux (2012) and the analysis has been extended to dynamic games by Weinstein and Yildiz (2011, 2016) and Penta (2012). The latter paper also allows for information types, and characterizes the strongest robust predictions in general information partitions with a product structure. Penta (2013) relaxes the richness condition in static games, and studies sufficient conditions for Weinstein and Yildiz's selection without richness; Chen et al. (2014) provide a full characterization. Aside from the shift from payoff to extensive form uncertainty, the present paper is the first to study the impact of higher order uncertainty with information types without richness.

The literature above as well as the present paper exploit infection arguments which date back to Rubinstein's (1989) email game, and which are also common in the contagion literature (e.g., Morris (2000), Steiner and Stewart (2008)), and in global games (e.g., Carlsson and Van Damme (1993), Morris and Shin (1998), Frankel, Morris and Pauzner (2003), Dasgupta, Steiner and Stewart (2010), Mathevet and Steiner (2013), etc.).

Related to extensive-form uncertainty, a few papers have studied games in which the timing of moves is endogenous (e.g., van Damme and Hurkens (1997, 1999, 2004)), games in which the information structure is endogenous (e.g., Solan and Yariv (2004)), or mechanism design settings in which players are uncertain over the mechanism the designer has committed to (Li and Peters (2017)). Zuazo-Garin (2017) introduces incomplete information over the information structure via type spaces, and studies sufficient conditions for the backward induction outcome. None of these papers, however, relax common knowledge assumptions in the sense that we do here, or in the literature on payoff uncertainty we just discussed.

An alternative approach to extensive-form robustness is that of Peters (2015), Ely and Doval (2016), Salcedo (2017) and Makris and Renou (2018), who seek to bound or characterize the distributions over outcomes which can be expected when the analyst only has limited information on the extensive form. These papers differ in the set of extensive forms considered in the analysis and in the equilibrium concepts they adopt. Ely and Doval (2016) and Makris and Renou (2018) also allow for payoff uncertainty. In all of them, however, it is maintained that the actual extensive form is common knowledge among the players.

The presence of types which face a hybrid of a static and dynamic optimization problem is reminiscent of aspects of the analysis in the literature on asynchronicity and reversibility of actions in games (e.g., Lagunoff and Matsui (1997), Kovac and Steiner (2013), Ambrus and Ishii (2015), Calcagno et al. (2014), Kamada and Kandori (2017a,b)). As discussed in Footnote 5, however, the sense in which this literature studies the impact of asynchronicity is very different from the one we pursue in Section 7.

## 2 Model

Consider a static two-player game $G^* = (A_i, u_i^*)_{i=1,2}$, where for any $i = 1, 2$, $A_i$ and $u_i^* : A_1 \times A_2 \to \mathbb{R}$ denote, respectively, $i$'s set of actions and payoff function, all assumed common knowledge. We let $A := A_1 \times A_2$ denote the set of (pure) action profiles, with typical element $a = (a_1, a_2)$, and let $NE^* \subseteq A$ denote the set of (pure) Nash Equilibrium profiles of $G^*$. Similar to the leading example in Section 1.1, we introduce extensive-form uncertainty by letting $\Omega = \{\omega^0, \omega^1, \omega^2\}$ denote the set of states of the world: state $\omega^0$ represents the state of the world in which the game is actually static; $\omega^i$ instead represents the state of the world in which the game has perfect information, with player $i$ moving first. (Some extensions of the model are discussed in Section 8.)

To avoid unnecessary technicalities, we maintain the following assumption on $G^*$:

**Assumption 1** *For each $i$ and for each $a_j \in A_j$, $\exists!\ a_i^*(a_j)$ s.t.* $\arg\max_{a_i \in A_i} u_i^*(a_i, a_j) = \{a_i^*(a_j)\}$ *and for each $A_i' \subseteq A_i$,* $\left|\arg\max_{a_i \in A_i'} u_i^*\left(a_i, a_j^*(a_i)\right)\right| = 1$.

This assumption, which is weaker than requiring that payoffs in $G^*$ are in generic position, ensures that backward induction is well-defined in both dynamic games associated to states $\omega^1$ and $\omega^2$, and for any subset of actions of the first mover. In the following, it will be useful to denote by $a^i = (a_1^i, a_2^i)$ the *backward induction outcome* in the game with perfect information in which player $i$ moves first (that is, the game in which $\omega^i$ is common knowledge).

**Information about the Extensive Form:** There are two possible pieces of 'hard information' for a player: either he knows he plays knowing the other's action (he is 'second', $\theta_i''$), or not (denoted by $\theta_i'$). Let $\Theta_i = \{\theta_i', \theta_i''\}$ denote the set of *information types*, generated by the information partition $\Omega$ with cells $\theta_i' = \{\omega^0, \omega^i\}$ and $\theta_i'' = \{\omega^j\}$. Hence, whereas the true state of the world is never *common knowledge* (although it may be common belief), it is always the case that it is *distributed knowledge*: $\theta = (\theta_1', \theta_2')$ if and only if $\omega = \omega^0$; $\theta = (\theta_1'', \theta_2')$ if and only if $\omega = \omega^2$; $\theta = (\theta_1', \theta_2'')$ if and only if $\omega = \omega^1$. Hence, if $\omega^0$ is the true state of the world, only types $\theta_i'$ are possible, but $i$ may still believe that $\theta_j = \theta_j''$, and this latter type would know that $\theta_i = \theta_i'$. In short, letting $\theta_i(\omega)$ denote the cell of $i$'s information partition which contains $\omega$, notice that $\theta_i(\omega) \cap \theta_j(\omega) = \{\omega\}$ for all $\omega \in \Omega$.

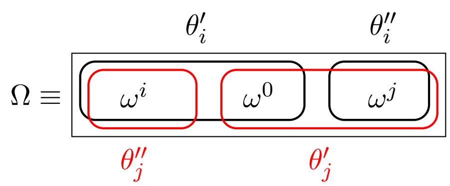


Figure 1: Uncertainty space and information partitions.

**Beliefs:** An *information-based type space* is a tuple $\mathcal{T} = (T_i, \hat{\theta}_i, \tau_i)_{i=1,2}$ s.t. $T_i$ is a compact and metrizable set of player $i$'s types, $\hat{\theta}_i : T_i \to \Theta_i$ assigns to each type his information about the extensive form, and beliefs $\tau_i : T_i \to \Delta(T_j)$ are continuous with respect to the weak* topology for $\Delta(T_j)$ and *consistent* with $i$'s information in the sense that a type's beliefs are concentrated on types of the opponents whose information is consistent with that of $t_i$ (formally: $\tau_i(t_i)[\{t_j : \hat{\theta}_i(t_i) \cap \hat{\theta}_j(t_j) \neq \emptyset\}] = 1$ for every $t_i \in T_i$ and $i$.)

As usual, any type in a (consistent) type space induces a belief hierarchy over $\Omega$. The consistency requirement restricts such hierarchies to be consistent with the type's information, but the construction is standard and so we leave it to the appendix. For any type $t_i$, and for any $k \in \mathbb{N}$, we let $\hat{\pi}_i^k(t_i)$ denote his $k$-th order beliefs. We let $T^*$ denote the *universal type space*, in which types coincide with belief-hierarchies (i.e. $t_i = (\hat{\theta}_i(t_i), \hat{\pi}_{i,1}(t_i), \hat{\pi}_{i,2}(t_i), ...)$ for each $t_i \in T_i^*$), as usual endowed with the product topology. Also, for any $\omega \in \Omega$, we let $t_i^{CB}(\omega)$ denote the type corresponding to common belief of $\omega$. For any consistent pair of types $t = (t_1, t_2)$, such that $\hat{\theta}_1(t_1) \cap \hat{\theta}_2(t_2) \neq \emptyset$, we let $\omega(t)$ denote the (unique) state of the world such that $\theta_i(\omega) = \hat{\theta}_i(t_i)$ for both $i = 1, 2$.

**Strategic Form:** Players' strategy sets depend on the state of the world:

$$S_i(\omega) = \begin{cases} A_i^{A_j} & \text{if } \omega = \omega^j \text{ and } j \neq i \\ A_i & \text{otherwise} \end{cases}$$

Note that $i$ knows his own strategy set at every state of the world (that is, $S_i : \Omega \to \{A_i\} \cup \{A_i^{A_j}\}$ as a function is measurable with respect to the information partition $\Theta_i$). Hence, with a slight abuse of notation, we can also write $S_i(t_i)$ to refer to $S_i(\omega)$ such that

$\omega \in \hat{\theta}_i(t_i)$. For any $\omega \in \Omega$, let $u_i(\cdot, \omega) : S(\omega) \to \mathbb{R}$ be such that:[6]

$$u_i(s_i, s_j, \omega) = \begin{cases} u_i^*(s_i, s_j) & \text{if } \omega = \omega^0, \\ u_i^*(s_i, s_j(s_i)) & \text{if } \omega = \omega^i, \\ u_i^*(s_i(s_j), s_j) & \text{if } \omega = \omega^j. \end{cases}$$

**Example 1** For instance, a baseline game $G^*$ such as the one in Figure 2a generates the two strategic forms, denoted $SF(\omega^1)$ and $SF(\omega^2)$, respectively depicted in Figures 2b and 2c. Hence, the dynamic games in states $\omega^1$ and $\omega^2$ have different normal forms, both of which are clearly distinct from $G^* \equiv SF(\omega^0)$. For instance, the backward induction outcome when $\omega^1$ is common knowledge, $(U, L)$, is not a Nash equilibrium of $G^*$, but of course it is an equilibrium outcome of $SF(\omega^1)$, induced by strategy profile $(U, LR)$ – the subgame-perfect equilibrium. Hence, our exercise cannot be described in terms of uncertainty over the extensive forms that may generate a given (common) strategic form. □

| | $L$ | $R$ |
|---|---|---|
| $U$ | 4 3 | 2 1 |
| $D$ | 5 1 | 3 2 |

(a) $G^* = SF(\omega^0)$

| | $LL$ | $LR$ | $RL$ | $RR$ |
|---|---|---|---|---|
| $U$ | 4 3 | 4 3 | 2 1 | 2 1 |
| $D$ | 5 1 | 3 2 | 5 1 | 3 2 |

(b) $SF(\omega^1)$

| | $L$ | $R$ |
|---|---|---|
| $UU$ | 4 3 | 2 1 |
| $UD$ | 4 3 | 3 2 |
| $DU$ | 5 1 | 2 1 |
| $DD$ | 5 1 | 3 2 |

(c) $SF(\omega^2)$

Figure 2: Normal form at each state of the world (Example 1).

## 3 Rationality and Common Belief in Rationality

We are interested in the behavioral implications of players' (sequential) rationality and common belief in (sequential) rationality in this setting. Under Assumption 1, these ideas can be conveniently expressed in the interim strategic form, by letting types who know they move second (i.e., $t_i$ such that $\hat{\theta}_i(t_i) = \theta_i''$) apply one round of deletion of *weakly* dominated strategies, followed by iterated strict dominance for all types. This boils down to applying Dekel and Fudenberg's (1990) $S^\infty W$ procedure to the interim strategic form of the game.[7]

Formally, fix a type space $\mathcal{T} = (T_i, \hat{\theta}_i, \tau_i)_{i=1,2}$. For any $i$ and $t_i$, we define the set of

[6] The following is well-defined because at $\omega^0$, $(s_i, s_j) \in S_i(\omega^0) \times S_j(\omega^0) = A_i \times A_j$, at $\omega^i$, $(s_i, s_j) \in S_i(\omega^i) \times S_j(\omega^i) = A_i \times A_j^{A_i}$ and at $\omega^j$, $(s_i, s_j) \in S_i(\omega^j) \times S_j(\omega^j) = A_i^{A_j} \times A_j$.

[7] In private value settings with payoff uncertainty, Dekel and Fudenberg's (1990) procedure in the interim normal form is equivalent to Penta's (2012) solution concept, and to Ben Porath's (1997) in complete information games. In two-stage games with complete and perfect information with no relevant ties, all these concepts yield the backward induction solution.

*feasible conjectures* and the *best responses to feasible conjectures*, respectively, as

$$C_i(t_i) := \{\mu \in \Delta(T_j \times S_j) : \mathrm{marg}_{T_j}\mu = \tau_i(t_i)\}, \text{ and for all } \mu \in C(t_i)$$
$$BR_i(\mu; t_i) := \mathrm{argmax}_{s_i \in S_i(t_i)} \int_{(t'_j, s'_j) \in T_j \times S_j} u_i(s_i, s'_j, \omega(t_i, t'_j)) d\mu.$$

For each $t_i$, we first set $R_i^0(t_i) = S_i(t_i)$ and $R_j^0 = \{(t_j, s_j) : s_j \in R_j^0(t_j)\}$. Then, in the first round, types who move second delete all weakly dominated strategies (to capture the idea that they are *sequentially* rational), whereas all other types – who either think they move first or simultaneously – only delete strictly dominated strategies: For each $t_i \in T_i$,

$$R_i^1(t_i) := \left\{ \hat{s}_i \in R_i^0(t_i) : \exists \mu \in \Delta\left(R_j^0\right) \cap C_i(t_i) \text{ s.t.: } \begin{array}{r} (i)\ \hat{s}_i \in BR_i(\mu; t_i) \\ (ii) \text{ if } \theta_i(t_i) = \theta''_i \text{ and } \mu(s_j, t_j) > 0, \\ \text{then } \mu\left(s'_j, t_j\right) > 0,\ \forall s'_j \in S_j(t_j) \end{array} \right\}.$$

For all subsequent rounds, all types perform iterated strict dominance: for all $k = 2, 3, ...$, having defined $R_j^{k-1} := \left\{(t_j, s_j) : s_j \in R_j^{k-1}(t_j)\right\}$, we let

$$R_i^k(t_i) := \left\{\hat{s}_i \in R_i^0(t_i) : \exists \mu \in \Delta\left(R_j^{k-1}\right) \cap C_i(t_i) \text{ s.t. } \hat{s}_i \in BR_i(\mu; t_i)\right\}, \text{ and}$$
$$R_i(t_i) := \bigcap\nolimits_{k \geq 0} R_i^k(t_i).$$

Hence, this solution concept is a hybrid of Interim Correlated Rationalizability (ICR, Dekel, Fudenberg and Morris, 2007) and Dekel and Fudenberg's (1990) $S^\infty W$ procedure. Arguments similar to Battigalli et al.'s (2011) therefore can be used to show that $R_i(t_i)$ characterizes the behavioral implications of (sequential) *Rationality and Common Belief of Rationality (RCBR)*, given type $t_i$.

**Example 2** Consider the following type space: For each player $i$, types are $T_i = \{t_i^1, t_i^0, t_i^2\}$, with information such that $\omega^x \in \hat{\theta}_i(t_i^x)$ for each $x = 0, 1, 2$. Beliefs are such that $\tau_i(t_i^x)[t_j^x] = 1$ if $x = 1, 2$, whereas $\tau_i\left(t_i^0\right)[t_j^0] = p$ and $\tau_i\left(t_i^0\right)[t_j^i] = 1 - p$. So, types $t_i^1$ and $t_i^2$ correspond to common belief that the game is dynamic, respectively with player 1 and player 2 as first mover. Type $t_i^0$ instead attaches probability $p$ to state $\omega^0$ and $(1-p)$ to state $\omega^i$. If $p = 1$, $t_i^0$ represents common belief in the static game. Next, consider the following game*:*

| | *Bach* | *Stravinsky* | *Chopin* |
|---|---|---|---|
| *Bach* | 4 2 | 0 0 | 0 0 |
| *Stravinsky* | 0 0 | 2 4 | 0 0 |
| *Chopin* | 0 0 | 0 0 | 1 1 |

First note that $S_i\left(t_i^i\right) = A_i$, whereas $S_j(t_j^i) = A_j^{A_i}$. Since no action is dominated for $t_i^i$, we have $R_i^1\left(t_i^i\right) = A_i$, whereas the non weakly dominated strategies for $t_j^i$ are $R_j^1(t_j^i) = \{s_j \in A_j^{A_i} : s_j(a_i) = a_i \text{ for all } a_i \in A_i\}$, which is in fact a singleton, whose element is denoted by

$a_j^*(\cdot)$. Given this, the only undominated action at the next round for $t_i^i$ is $R_i^2\left(t_i^i\right) = \left\{a_i^i\right\}$, and hence the only outcome consistent with $R\left(t^i\right)$ is $a^i = \left(a_i^i, a_j^*\left(a^i\right)\right)$. This is precisely the backward induction solution, only obtained applying Dekel and Fudenberg's procedure to the strategic form of the two-stage game, if types commonly believe $\omega^i$. If $p = 1$, (i.e., if $t^0$ represents common belief that the game is static) it is also easy to check that $R\left(t^0\right) = A$ coincides with standard (static) rationalizability (Bernheim (1984) and Pearce (1984)). $\square$

Many of our results will be concerned with the behavior of the $R(\cdot)$ correspondence (i.e., RCBR) around $t^{CB}\left(\omega^0\right)$ and $t^{CB}\left(\omega^i\right)$, respectively the benchmarks in which players commonly believe that the game is static, or that $i$ moves first. As shown by Example 2, $R(\cdot)$ coincides with standard (static) rationalizability for the former (types $t^0$ with $p = 1$ in the Example), and with the backward induction solution for the latter (types $t^i$ in the Example). The behavior of $R(\cdot)$ when such belief hierarchies are perturbed, however, will in general depend on its solutions for other belief hierarchies, including for instance those in which players are uncertain over whether the game is static or not.

**Example 3 (Ex.2, continued.)** Consider again the game and type space in Example 2. If $p \in (0,1)$, types $t_i^0$ are uncertain as to whether their action will be osberved by the opponent. In this case, type $t_i^0$ attaches probability $p$ to playing a static game against type $t_j^0$, and probability $(1-p)$ to playing the dynamic game in which he goes first against type $t_j^i$. Then, it is easy to check that, for $i = 1, 2$, $R_i^1(t_i^j) = \{a_i^*(\cdot)\}$ and $R_i^1\left(t_i^0\right) = R_i^1\left(t_i^i\right) = A_i$. At the second round, types $t_i^i$ assign probability one to the opponent's type $t_j^i$, who plays $a_j^*(\cdot)$, and hence play the backward induction action $a_i^i$: $R_i^2\left(t_i^i\right) = R_i\left(t_i^i\right) = \left\{a_i^i\right\}$, $R_i^1(t_i^j) = R_i(t_i^j) = \{a_i^*(\cdot)\}$. For type $t_i^0$ instead the problem is more complicated: with probability $(1-p)$ he thinks the game is dynamic, and that he faces type $t_i^j$ who plays $a_j^*(\cdot)$; with complementary probability instead he faces $t_j^0$, for whom $R_j^1(t_j^0) = A_j$, and so he will have to form conjectures $\mu \in \Delta(A_j)$ over that type's behavior. The resulting optimization problem for type $t_i^0$, with conjectures $\mu$ over $t_j^0$'s action, is therefore to choose $a_i' \in A_i$ that maximizes the following expected payoff:

$$EU\left(a_i'; p, \mu\right) := \left(p \cdot \sum_{a_j \in A_j} \mu[a_j] \cdot u_i^*(a_i', a_j) + (1-p) \cdot u_i^*(a_i', a_j^*(a_i'))\right). \tag{1}$$

Hence, $R_i^2\left(t_i^0\right) = \{a_i \in A_i : \exists \mu \in \Delta(A_j) \text{ s.t. } a_i \in \arg\max_{a_i' \in A_i} EU\left(a_i', \mu; t_i^0\right)\}$. As a function of $p$, this set is equal to

$$R_i^2\left(t_i^0\right) = R_i\left(t_i^0\right) = \begin{cases} A_i & \text{if } p \geq 3/4, \\ \{Bach, Stravinsky\} & \textit{if } p \in [1/2, 3/4), \\ \{Bach\} & \textit{if } p < 1/2. \end{cases}$$

Intuitively, if $p$ is sufficiently high (i.e., $p \geq 3/4$), type $t_i^0$ attaches a sufficiently high probability to the game being static, that all actions can be justified on the grounds of RCBR. Hence, for high $p$, $R_i\left(t_i^0\right)$ coincides with standard (static) rationalizability when players commonly believe that the game is static. At the opposite extreme, when $p < 1/2$, then $t_i^0$ attaches sufficiently high probability to being the first mover in a dynamic game with perfect information, that RCBR essentially delivers the backward induction solution of the dynamic game in which $i$ moves first. For intermediate levels of $p$ (i.e., when $p \in [1/2, 3/4)$), both *Bach* and *Stravinsky* are consistent with RCBR, but not *Chopin*.

To see this, let $\mu^S$ (respectively, $\mu^B$ and $\mu^C$) denote $t_1^0$'s conjectures that attach probability 1 to type $t_2^0$ playing *Stravinsky* (resp., *Bach* and *Chopin*), which gives the best shot at *Stravinsky* (resp. *Bach* and *Chopin*) being optimal for type $t_1^0$. Then:

$$
\begin{aligned}
EU_1\left(B; p, \mu^B\right) &= p \cdot 4 + (1-p)\, 4 = 4 \\
EU_1\left(B; p, \mu^C\right) &= EU_1\left(B; p, \mu^S\right) = p \cdot 0 + (1-p)\, 4 = 4 - 4p \\
EU_1\left(S; p, \mu^S\right) &= p \cdot 2 + (1-p)\, 2 = 2 \\
EU_1\left(S; p, \mu^C\right) &= EU_1\left(S; p, \mu^B\right) = p \cdot 0 + (1-p)\, 2 = 2 - 2p \\
EU_1\left(C; p, \mu^C\right) &= p \cdot 1 + (1-p)\, 1 = 1 \\
EU_1\left(C; p, \mu^S\right) &= EU_1\left(C; p, \mu^B\right) = p \cdot 0 + (1-p)\, 1 = 1 - p
\end{aligned}
$$

Note that $4 - 4p > 2$ whenever $p < 1/2$, and hence *Bach* dominates *Stravinsky* for $p < 1/2$. Similarly, $4 - 4p > 1$ whenever $p < 3/4$, and hence *Bach* dominates *Chopin* for $p < 3/4$. It follows that *Chopin* is never rationalizable when $p < 3/4$, and *Stravinsky* is never rationalizable when $p < 1/2$. To see that both *Bach* and *Stravinsky* are rationalizable when $p \in [1/2, 3/4)$, note that in this range $EU_1\left(S; p, \mu^S\right) \geq EU_1\left(B; p, \mu^S\right)$ and $EU_1\left(B; p, \mu^B\right) > EU_1\left(S; p, \mu^B\right)$: hence, if $p \in [1/2, 3/4)$, both *Bach* and *Stravinsky* can be justified by some conjectures about $t_2^0$'s behavior, for instance by conjectures such as $\mu^B$ and $\mu^S$, respectively. A symmetric argument shows the result for type $t_2^0$. □

The combination of static and dynamic best-responses illustrated in the previous example, and formalized by an objective function such as (1), will play a central role in the analysis that follows. As we will explain, this non-standard feature of our analysis will be one of the main determinants of the difference between our results and those in Weinstein and Yildiz (2007, 2011, 2013, 2016) and Penta (2012, 2013): in those papers, which consider payoff uncertainty, the underlying game structure is either static or dynamic, and there is no uncertainty over the extensive form. Hence, player's objective functions are either 'fully static' or 'fully dynamic'. Here, in constrast, the cases in which $p \in (0, 1)$ in equation (1) are central to the analysis, and yield significant differences in the behavior of players' best-response correspondences, which crucially affect our general results in Section 4.

Before moving on to our characterization of the robust predictions under extensive-form uncertainty, however, it is useful to discuss two important robustness properties of our solution concept, $R_i$ – namely, RCBR – in the present context.

**Lemma 1 (Type space invariance)** *For any two type spaces* $\mathcal{T}$ *and* $\tilde{\mathcal{T}}$*, if* $t_i \in T_i$ *and* $\tilde{t}_i \in \tilde{T}_i$ *are such that* $(\hat{\theta}_i(t_i), \hat{\pi}_i(t_i)) = (\hat{\theta}_i(\tilde{t}_i), \hat{\pi}_i(\tilde{t}_i))$*, then* $R_i(t_i) = R_i(\tilde{t}_i)$.

This result ensures that the predictions of $R_i(\cdot)$ only depend on a type's information and belief hierarchy, not on the particular type space used to represent it. This property, which may itself be regarded as a robustness property of the solution concept (cf. Penta, 2012), is convenient in this case because it enables us to study $R(\cdot)$ as a correspondence on the universal type space, $R_i : T_i^* \rightrightarrows S_i$, without keeping track of the particular type space. This is a standard property for solution concepts that allow for correlated conjectures, such as Dekel, Fudenberg and Morris' (2006, 2007) interim correlated rationalizability (ICR) and Penta's (2012) interim sequential rationalizability (ISR), and it plays the same role in the analyses of Weinstein and Yildiz (2007) and Penta (2012). In those papers, the underlying uncertainty is on players' payoffs, whereas here the uncertainty is over the extensive form, but the logic of the proof is essentially the same.[8]

The next result states yet another standard property for $R_i$, envisioned as a correspondence on the universal type space: upper hemicontinuity.

**Lemma 2 (Upper-hemicontinuity)** *$R_i : T_i^* \rightrightarrows S_i$ is an* upper-hemicontinuous *(u.h.c.) correspondence: if $t_i^\nu \to t_i$ and $s_i \in R_i(t_i^v)$ for all $\nu$, then $s_i \in R_i(t_i)$.*

This result shows that, similar to ICR and ISR on the universal type space generated by an underlying space of payoff uncertainty, $R_i$ is upper hemicontinuous on the universal type space generated by $\Omega$. This is a robustness property in that it ensures that anything that is ruled out by $R_i$ for some type $t_i \in T_i^*$, would also be ruled out for all types in a neighborhood of $t_i$. Upper hemincontinuity plays an important role in the above mentioned literature. For instance, the unrefinability results in Weinstein and Yildiz (respectively, Penta (2012)) can be summarized by saying that ICR (respectively, ISR) is the strongest u.h.c. solution concept among its refinements.

As we will show shortly, however, whereas $R_i$ is u.h.c. on $T_i^*$, with the extensive-form uncertainty we consider here it will not be the *strongest* u.h.c. solution concept: a proper refinement of $R_i$ is also u.h.c. Hence, the 'strongest robust' predictions under extensive-form uncertainty support a proper refinement of rationalizability, which we characterize next.

# 4 Robust Predictions: Characterization

In this section we characterize the strongest predictions consistent with RCBR that are robust to higher order uncertainty over the extensive form. We begin by constructing a set of actions, $\mathcal{B}_i \subseteq A_i$, which consists of all actions that can be uniquely rationalized for some type in the universal type space. The intuitive idea behind this construction is best understood thinking about our leading example in Section 1.1. There, an 'infection argument' showed that the uniqueness of the backward induction solution for types that commonly believe in $\omega^i$ propagates to types sharing $n$ levels of mutual belief in $\omega^0$ through a chain of unique best replies. As discussed in Section 1.2, this is a standard argument in the literature.

[8] Type space invariance is typically *not* satisfied by solution concepts that involve assumptions of conditional independence on players' conjectures, such as Bayes-Nash equilibrium or interim independent rationalizability (Ely and Pęski(2006)), which may be affected by the possibility of 'redundant types' (types that induce the same belief-hierarchy). In contrast, solution concepts such as ICR, ISR, or $R$ above, which already allow all possible correlations in players' conjectures, are unaffected by the extra-correlation which may be associated with redundant types, and hence they only depend on the belief hierarchies (cf. Dekel et al. (2007)).

In general, these arguments have two main ingredients: $(i)$ the *seeds* of the infection, and $(ii)$ a chain of *strict best responses*, which spreads the infection to other types. In Weinstein and Yildiz (2007), for instance, best responses are the standard ones that define rationality in static games, whereas a 'richness condition' ensures that any action is dominant at some state, and hence the infection can start from many 'seeds', one for every action of every player.[9] Due to the nature of the uncertainty we consider, both elements will differ from Weinstein and Yildiz's in our analysis: first, only the backward induction outcomes can serve as seeds (see the examples in Section 1.1); second, best responses must account for the 'hybrid' problems that extensive-form uncertainty may generate (see Example 3 and equation (1)).[10] The set $\mathcal{B}_i$ is defined recursively, based precisely on these two elements. Formally: for each $i$, let $\mathcal{B}_i := \bigcup_{k\geq 1} \mathcal{B}_i^k$, where $\mathcal{B}_i^1 := \{a_i^i\}$ (the 'seeds') and recursively, for $k \geq 1$,

$$\mathcal{B}_i^{k+1} := \mathcal{B}_i^k \cup \left\{ a_i \in A_i : \begin{array}{c} \exists \mu^i \in \Delta(\mathcal{B}_j^k), \exists p \in [0,1] \text{ s.t.:} \\ \{a_i\} = \arg\max\limits_{a_i' \in A_i} \left( p \cdot \sum\limits_{a_j \in A_j} \mu^i[a_j] \cdot u_i^*(a_i', a_j) + (1-p) \cdot u_i^*(a_i', a_j^*(a_i')) \right) \end{array} \right\}.$$

Since $A$ is finite, there exists some $m < \infty$ such that $\mathcal{B}_i^m = \mathcal{B}_i$ for all $i$. If $p = 1$ in the definition of $\mathcal{B}_i^{k+1}$, then $\mathcal{B}_i^{k+1}$ contains the strict best replies in the static game to conjectures concentrated on $\mathcal{B}_j^k$. The case $p < 1$ instead corresponds to a situation in which $i$ attaches probability $(1-p)$ to player $i$ observing his choice $a_i$, and hence respond by chosing $a_j^*(a_i)$. Hence, as $p$ varies between 0 and 1, $\mathcal{B}_i^{k+1}$ may also contain actions that are *not* a static best response to conjectures concentrated in $\mathcal{B}_j^k$. The following example illustrates the point:

**Example 4** Consider the following game, where $x \in [0,1]$:

| | $L$ | | $C$ | | $R$ | |
|---|---|---|---|---|---|---|
| $U$ | 4 | 2 | 0 | 0 | 0 | 0 |
| $M$ | 0 | 0 | 2 | 4 | 0 | 0 |
| $D$ | 0 | 0 | $x$ | 0 | 3 | 3 |

Then, $a^1 = (U, L)$ and $a^2 = (C, R)$, and hence $B_1^1 = \{U\}$, $B_2^1 = \{C\}$. Since $M$ (respectively $L$) is a unique best response to $C$ (resp. $U$), it follows that $M \in \mathcal{B}_1^2$ (resp., $L \in \mathcal{B}_2^2$). Moreover, it can be checked that no other actions are a best response for any $p \in [0,1]$, hence $\mathcal{B}_1^2 = \{U, M\}$, $\mathcal{B}_2^1 = \{C, L\}$. At the third iteration, suppose that $\hat{\mu}^1$ attaches probability one

[9] Similarly, the analysis of dynamic games in Penta (2012) can be thought of as maintaining seeds for all actions, but accounting for sequential rationality considerations in the best replies. Penta (2013) and Chen, Takahashi and Xiong. (2014) instead keep standard static rationality, but relax the richness assumption and hence allow for a smaller set of seeds.

[10] Another difference is that types in our setting also incorporate information. With the exception of Penta (2012), which allowed for information partitions with a product structure, all other papers cited in the previous footnote (as well as Weinstein and Yildiz (2007, 2011, 2013, 2016)) maintain that types have no information. Thus, due to the combined presence of an information partition (in this case without a product structure) and the absence of a richness condition (which had only been relaxed by in static games and with no information, see previous footnote), the current analysis cannot be cast within any of the existing frameworks.

to $C \in \mathcal{B}_2^2$, and let $p \in [0,1]$. Then, the expected payoffs from player 1's actions are:

$$EU_1\left(U; p, \hat{\mu}^1\right) = p \cdot 0 + (1-p)\,4 = 4 - 4p$$
$$EU_1\left(M; p, \hat{\mu}^1\right) = p \cdot 2 + (1-p)\,2 = 2$$
$$EU_1\left(D; p, \hat{\mu}^1\right) = p \cdot x + (1-p)\,3 = 3 - (3-x)\,p$$

If $x = 1$, $D$ is the only maximizer when $p \in (1/6, 1/2)$, and hence $D \in \mathcal{B}_1^3$. It follows that, if $x = 1$, $\mathcal{B}_i = A_i$ for both players in this game. If instead $x = 0$, then it is easy to check that $\mathcal{B}_1 = \{U, M\}$ and $\mathcal{B}_2 = \{L, C\}$. □

We introduce next the solution concept which we will show characterizes the robust predictions of RCBR under extensive-form uncertainty. We will denote such 'robust predictions' by the correspondence $RP_i : T_i^* \rightrightarrows A_i$.

Intuitively, $RP_i$ is obtained by applying the same iterated deletion procedure as $R_i$, but starting from the set $\mathcal{B}$ instead of $A$. This way, $RP_i$ yields – for types $t_i$ s.t. $\theta_i(t_i) = \theta_i'$ – the largest subset of $\mathcal{B}$ with the best-reply property, where the latter is defined taking into account the beliefs consistent with type $t_i$ (that is, regarding the types he faces, as well as the induced probability that the opponent observes his action). But since, under the maintained Assumption 1, strategy $a^*(\cdot)$ is the only strategy that is not weakly dominated for all types that move second (see Lemma 7 in Appendix B), it is convenient to initialize the procedure directly from this point. Formally: for each $i$ and $t_i$, let

$$RP_i^0(t_i) := \begin{cases} \mathcal{B}_i & \text{if } \theta_i(t_i) = \theta_i', \\ \{a^*(\cdot)\} & \text{otherwise.} \end{cases}$$
$$RP_i^0 := \{(t_i, s_i) \in T_i^* \times S_i : s_i \in RP_i^1(t_i)\}.$$

For all the subsequent rounds, all types perform iterated strict dominance: Inductively, for all $k = 1, 2, ...$, having defined $RP_j^{k-1} = \{(t_j, s_j) : s_j \in RP_j^{k-1}(t_j)\}$, we let

$$RP_i^k(t_i) = \left\{\hat{s}_i \in RP_i^0(t_i) : \exists \mu \in \Delta(RP_j^{k-1}) \cap C_i(t_i) \text{ s.t. } \hat{s}_i \in BR_i(\mu; t_i)\right\}$$
$$\text{and } RP_i(t_i) := \bigcap\nolimits_{k \geq 0} RP_i^k(t_i).$$

Obviously, $RP_i$ and $R_i$ coincide if $\mathcal{B} = A$, but in general $RP_i(t_i) \subseteq R_i(t_i) \cap \mathcal{B}_i$ for all $t_i$ s.t. $\hat{\theta}_i(t_i) = \theta_i'$, whereas $RP_i(t_i) = R_i(t_i) = \{a^*(\cdot)\}$ for all $t_i$ s.t. $\hat{\theta}_i(t_i) = \theta_i''$. Hence, $RP_i$ is a refinement of $R_i$. In the static benchmark, $RP_i$ can be computed easily, as the largest subset of $\mathcal{B}$ that has a standard best-response property (see Lemma 7 in Appendix.)

The next theorem provides the main results of the paper, and formalizes the sense in which $RP_i$ characterizes the strongest robust predictions consistent with RCBR under extensive-form uncertainty, and that both $RP_i$ and $R_i$ are generically unique and they coincide:

**Theorem 1 (Robust Predictions)** *For any player $i$ the following three properties hold:*

(*i*) *For any $k \in \mathbb{N}$, if $(\theta_i(t_i), \hat{\pi}_{i,k}(t_i)) = (\theta_i(\check{t}_i), \hat{\pi}_{i,k}(\check{t}_i))$, then $RP_i^k(t_i) = RP_i^k(\check{t}_i)$.*

(*ii*) $RP_i : T_i^* \rightrightarrows A_i$ *is non-empty valued and upper hemicontinuous.*

(*iii*) *For any finite type* $t_i$ *and any strategy* $s_i \in RP_i(t_i)$ *there exists a sequence of finite types* $\{t_i^\nu\}_{\nu \in \mathbb{N}}$ *with limit* $t_i$ *and such that* $R_i(t_i^\nu) = RP_i(t_i^\nu) = \{s_i\}$ *for any* $\nu \in \mathbb{N}$.

The first part of Theorem 1 states that, for every $k$, $RP_i^k$ only depends on the $k$ lower order beliefs. Part (*ii*) ensures that the predictions of $RP_i(\cdot)$ are robust to higher-order uncertainty on the extensive form: anything that is ruled out by $RP_i$ for a particular type $t_i$ would still be ruled out for all types in a neighborhood of $t_i$. The third part states that, for any finite type $t_i$, any strategy $s_i \in RP_i(t_i)$ is uniquely selected by both $R_i(\cdot)$ and $RP_i(\cdot)$ for some finite type arbitrarily close to $t_i$. This has a few important implications: (*i*) first, $RP_i(\cdot)$ characterizes the *strongest* predictions that are robust, since no refinement of $RP_i(\cdot)$ is upper hemicontinuous; (*ii*) second, the two solution concepts ($R_i$ and $RP_i$) generically coincide on the universal type space, and deliver the same unique prediction – hence, not only $RP_i(\cdot)$ is a strongest upper hemicontinuous refinement of $R_i(\cdot)$, but in fact it characterizes the predictions of $R_i$ which do not depend on the fine details of the infinite belief hierarchies (what we call the 'robust predictions' of RCBR); (*iii*) finally, since $RP_i(\cdot)$ is upper hemicontinuous, the 'nearby uniqueness' result only holds for the strategies in $RP_i(t_i)$, not for those in $R_i(t_i) \backslash RP_i(t_i)$. We summarize this discussion in the following corollaries:

**Corollary 1** *No proper refinement of* $RP_i$ *is upper hemicontinuous on* $T_i^*$.

**Corollary 2** $R_i$ coincides with $RP_i$ *and is* single-valued *over an open and dense set of types in the universal type space.*

**Corollary 3** *For any* $t_i$, *if there exists a sequence* $\{t_i^\nu\}_{\nu \in \mathbb{N}} \subseteq T_i^*$ *with limit* $t_i$ *such that* $R_i(t_i^\nu) = \{s_i\}$ *for all* $\nu \in \mathbb{N}$, *then* $s_i \in RP_i(t_i)$

Hence, while there is a clear formal similarity between Theorem 1 and the famous result of Weinstein and Yildiz (2007), the implications of Theorem 1 (and particularly Corollaries 1-3) are very different: higher order uncertainty over the extensive form supports a *robust refinement* of RCBR, characterized precisely by $RP_i(\cdot)$. Clearly, in games in which $\mathcal{B} = A$ (e.g., in a standard Battle of the Sexes), $R_i\left(t_i^{CB}\left(\omega^0\right)\right) = RP_i\left(t_i^{CB}\left(\omega^0\right)\right)$, and hence the results have the same implications, conceptually. But in some cases the difference can be especially sharp.

**Example 5** Consider the following game:

| | $L$ | $C$ | $R$ |
|---|---|---|---|
| $U$ | 4  2 | 0  0 | 0  0 |
| $M$ | 6  0 | 2  4 | 0  0 |
| $D$ | 0  0 | 0  0 | 3  3 |

Note that action $U$ is dominated by $M$ for player 1, whereas $L$ is only a best response to $U$. It follows that the set of rationalizable action profiles for this game, if players commonly believe that neither player observes the other's action, is $R(t^{CB}(\omega^0)) = \{M, D\} \times \{C, R\}$. Moreover, the backward induction outcomes are $a^1 = (U, L)$ and $a^2 = (M, C)$, and it is easy to check that $\mathcal{B} = \{U, M\} \times \{L, C\}$. It follows that in this case $RP(t^{CB}(\omega^0)) = R(t^{CB}(\omega^0)) \cap \mathcal{B} = \{(M, C)\}$. Hence, $(M, C)$ is the unique robust prediction for this static game. □

The result that $R_i$ and $RP_i$ generically coincide (Corollary 2) is particularly relevant from a conceptual viewpoint: Suppose that, for purely epistemic considerations (or other *a priori* reasons), we had decided to only care about the predictions generated by RCBR, except that we do not want to rely on the fine details of the infinite belief hierarchies on the extensive form. Then, Corollary 2 implies that whereas RCBR may deliver less sharp predictions than $RP(\cdot)$ for non-generic types (such as $t^{CB}(\omega^0)$ in the example, where RCBR only rules out $U$ and $L$), it would still be unique and coincide with $RP_i(\cdot)$ generically on the universal type space. In this sense, $RP_i(\cdot)$ characterizes the 'robust predictions' of RCBR.

Note that, in Example 5, not only are the robust predictions particularly sharp, but they also imply that arbitrarily small higher order uncertainty on the extensive form ensures that players achieve equilibrium coordination under common belief in rationality, i.e. without imposing correctness of beliefs. In Section 6 we will consider important cases in which the robust predictions take this especially strong form, and so equilibrium coordination arises purely from individual reasoning (*eductive coordination*). Section 5 instead explores other classes of games, in which Theorem 1 has particularly strong implications, which may or may not lead to eductive coordination. We conclude this section with one more example, which illustrates an obvious but interesting non-monotonicity of the robust predictions:

**Example 6** The following game is obtained from Ex.5 by dropping action $U$:

| | $L$ | $C$ | $R$ |
|---|---|---|---|
| $M$ | 6 0 | 2 4 | 1 0 |
| $D$ | 0 0 | 0 0 | 3 3 |

As in the previous example, the rationalizable set under common belief in $\omega^0$ is $R\left(t^{CB}\left(\omega^0\right)\right) = \{M, D\} \times \{C, R\}$. However, in this case the backward induction outcomes are $a^1 = (D, R)$ and $a^2 = (M, C)$. It follows that $\mathcal{B} = \{M, D\} \times \{C, R\}$ and it is easy to show that $RP\left(t^{CB}\left(\omega^0\right)\right) = \{M, D\} \times \{C, R\}$. Hence, eliminating actions from a game may enlarge the set of robust predictions. □

# 5 Applications to Robust Refinements

In this section we explore some implications of Theorem 1 in *agreement* and *coordination games* (Section 5.1) and in *strictly competitive games* (Section 5.2). As we will show, the robust predictions of RCBR select actions associated to efficient Nash Equilibria in the former, and uniquely selects the maxmin solution in the latter. Section 5.3 discusses some extensions.

### 5.1 Agreement and Coordination Games

We introduce next a class of games that generalizes the example in Section 1.1.

**Definition 1 (Agreement Games)** *$G^*$ is an* agreement game *if $NE^* \neq \emptyset$ and for any player $i$, any $a \in NE^*$ and any $a' \notin NE^*$, $u_i^*(a) > u_i^*(a')$. If, in addition, for any player $i$ it holds that $u_i^*(a') = u_i^*(a'')$ for all $a', a'' \notin NE^*$, then we say that $G^*$ has* constant disagreement payoffs*.*

Hence, in an *agreement game* both players strictly prefer any Nash equilibrium to any non-Nash equilibrium outcome. Note that, for instance, the game in Section 1.1 satisfies both properties in Definition 1. The next result shows that the insight from that example generalize to all such games:

**Proposition 1** *In agreement games with constant disagreement payoffs, RCBR selects one action from the backward induction profiles generically in the universal type space; i.e., if $G^*$ is an agreement game with constant disagreement payoffs, there exists an open and dense set $T_i' \subseteq T_i^*$ such that, for any type $t_i \in T_i'$, $R_i(t_i)$ is unique and such that $R_i(t_i) = RP_i(t_i) \in \{\{a_i^i\}, \{a_i^j\}\}$ if $\hat{\theta}_i(t_i) = \theta_i'$ and $R_i(t_i) = \{a^*(\cdot)\}$ if $\hat{\theta}_i(t_i) = \theta_i''$.*

**Corollary 4** *In an agreement game with constant disagreement payoffs, actions associated to inefficient Nash equilibria are generically ruled out by RCBR.*

It is not difficult to show that, in agreement games with constant disagreement payoffs, the benchmark 'static' types $t_i^{CB}(\omega^0)$ are such that $RP_i(t_i^{CB}(\omega^0)) = \{a_i^i, a_i^j\}$, and hence Theorem 1 implies that both $RP_i$ and $R_i$ *uniquely* select one of the backward induction actions in a neighborhood of $t_i^{CB}(\omega^0)$. The statement of Proposition 1, however, is stronger than that, and it refers to the generic predictions of RCBR (i.e., solution concept $R_i$), not to $RP_i(\cdot)$. Similar to the earlier discussion on the significance of Corollary 2, this emphasizes the idea that, if one cares about the predictions generated by RCBR which do not depend on the fine details of the infinite belief hierarchies on the extensive form, then Proposition 1 says that whereas $R_i$ may deliver less sharp predictions for non-generic hierarchies of beliefs, it would still be unique and select from the set $\{a_i^i, a_j^i\}$ *generically* in the universal type space.

Hence, for instance, whereas actions associated to inefficient Nash equilibria of $G^*$ are consistent with RCBR when $\omega^0$ is common belief, generically they are not: as soon as higher order uncertainty over the extensive form is considered, RCBR generically selects one of the two backward induction outcomes, and hence actions associated to inefficient equilibria of $G^*$ are consistent with RCBR only for non generic types (such as $t^{CB}(\omega^0)$).

The proof of Proposition 1 is simple: it follows from the observation that, in agreement games with constant disagreement payoffs, $\mathcal{B}_i = \{a_i^i, a_i^j\}$, and that by definition $RP_i(t_i) \subseteq \mathcal{B}_i$ whenever $\theta_i(t_i) = \theta_i'$. This, together with the fact that $RP_i(\cdot) = R_i(\cdot)$ generically on $T^*$ (Corollary 2), implies the result. Note that, because of Theorem 1, Proposition 1 could be equivalently stated for $R_i$ and $RP_i$. That is because not only $RP_i$ is a strongest u.h.c. refinement of $R_i$, but in fact it characterizes its robust predictions.

To better understand the class of games in the statement of Proposition 1, it is useful to discuss their relationship with the more common notion of *coordination games*. We are aware of no agreed upon textbook definition of coordination game, but the following definition may serve as a relatively uncontroversial common ground:

**Definition 2 (Coordination Games)** *A* coordination game *is a game in which every profile in which players choose the same or corresponding (pure) strategies is a strict Nash-equilibrium. If, in addition, for any player $i$ it holds that $u_i^*(a') = u_i^*(a'')$ for all $a', a'' \notin NE^*$, then we say that $G^*$ has* constant non-equilibrium payoffs.[11]

**Lemma 3** *A coordination game with constant non-equilibrium payoffs is an agreement game with constant disagreement payoffs. Hence, the results in Proposition 1 and Corollary 4 also apply to coordination games with constant non-equilibrium payoffs.*

For later reference, it is also useful to introduce the more general notion of *Nash commitment games*, in which both backward induction outcomes, $a^1$ and $a^2$, are Nash equilibria of the baseline game $G^*$:[12]

**Definition 3 (Nash-Commitment)** *$G^*$ is a* Nash commitment game *if $\{a^1, a^2\} \subseteq NE^*$.*

**Lemma 4** *If $G^*$ is an agreement or a coordination game, it is a Nash commitment game.*

### 5.2 Strictly Competitive Games

The previous section applied Theorem 1 to agreement and coordination games, which represent situations in which players' incentives are somewhat aligned. In this section we consider zero-sum games, also referred to as 'strictly competitive' games (e.g., Luce and Raiffa (1957)), which instead represent the archetypical model of situations of conflict.

**Definition 4** *$G^*$ is* strictly competitive *(or* zero-sum*) if, for all $a \in A$, $u_i^*(a) = -u_j^*(a)$.*

We begin by considering games with pure equilibria. (The no pure equilibrium case is discussed in Section 5.3.) In these settings, the celebrated maxmin solution of von Neumann and Morgenstern (1944, 1947) provides a very sharp prediction, which is sometimes at odds with the indeterminacy of RCBR. We will show that this tension disappears once higher order uncertainty over the extensive form is introduced.

**Example 7** Consider the following zero-sum game:

[11] To clarify the meaning of 'corresponding' in this definition, a coordination game is such that there exists an ordering of players' actions, $\{a_i(1), ..., a_i(n^*)\} = A_i$, such that all profiles of the form $(a_i(n), a_j(n))$ are Nash equilibria. *Coordination games with constant non-equilibrium payoffs* are sometimes called 'Unanimity Games' (Harsanyi (1981), see also Kalai and Samet (1984)). As shown by Lemma 3, *agreement games with constant disagreement payoffs* are a generalization of unanimity games.

[12] We thank Drew Fudenberg for suggesting the term.

| | $L$ | | $C$ | | $R$ | |
|---|---|---|---|---|---|---|
| $U$ | $2$ | $-2$ | $-1$ | $1$ | $-2$ | $2$ |
| $M$ | $1$ | $-1$ | $0$ | $0$ | $1$ | $-1$ |
| $D$ | $-2$ | $2$ | $-1$ | $1$ | $2$ | $-2$ |

In this game, player 1's 'security levels' (i.e., minimum payoffs) from actions $U, M, D$ are, respectively $(-1, 0, -1)$, and hence 1's maxmin action is $M$. Similary, 2's security levels from $L, C, R$ are, respectively, $(-1, 0, -1)$, and hence player 2's maxmin action is $C$. Furthermore, $M$ and $C$ are mutual best responses; hence $(M, C)$ is the only maxmin profile for this game, and the resulting payoff of 0 is the value of this game. It is also easy to see that everything is rationalizable in this game, since $U$ is justified by $L$, which is justified by $D$, which in turn is justified by $R$, which is a best-reponse to $U$. However, it is easy to see that $a^1 = a^2 = (M, C)$, and hence $\mathcal{B} = RP(t) = \{(M, C)\}$ for all $t \in T^*$. The maxmin solution therefore is the unique robust prediction in this game, and hence it is the only prediction consistent with RCBR generically in the universal type space. □

We show next that the example's insights are fully general:

**Proposition 2** *If $G^*$ is a zero-sum game with a pure strategy equilibrium, $a^* \in NE^*$, then: (i) $a^* = a^1 = a^2$, (ii) $\mathcal{B} = \{a^*\}$, and (iii) $RP(t) = \{a^*\}$ for all $t \in T^*$. Hence, the maxmin solution $a^*$ is also the unique prediction of RCBR generically on the universal type space.*[13]

One interesting aspect of this result is that it bridges a gap between RCBR and the maxmin solution which has long been discussed in the literature (e.g., Luce and Raiffa (1957), Schelling (1960)). To illustrate the point, we adapt arguments from Luce and Raiffa (1957) to a variation of the game in Example 7:[14]

**Example 8** Consider the following game, in which $\varepsilon > 0$:

| | $L$ | | $C$ | | $R$ | |
|---|---|---|---|---|---|---|
| $U$ | 1M | −1M | $-\varepsilon$ | $\varepsilon$ | $-2\varepsilon$ | $2\varepsilon$ |
| $M$ | $\varepsilon$ | $-\varepsilon$ | $0$ | $0$ | $\varepsilon$ | $-\varepsilon$ |
| $D$ | $-2\varepsilon$ | $2\varepsilon$ | $-\varepsilon$ | $\varepsilon$ | $2\varepsilon$ | $-2\varepsilon$ |

As in Example 7, everything is rationalizable in this game, $(M, C) := a^*$ is the maxmin solution, and $\mathcal{B} = \{(M, C)\}$. In Luce and Raiffa's words, choice [$M$] has two properties for player 1: "($i$) It maximizes player 1's security level; ($ii$) it is the best counterchoice against

[13]In zero-sum games, Assumption 1 rules out the possibility of multiple (payoff equivalent) equilibria.

[14]Luce and Raiffa's original argument refers to a game that violates Assumption 1, which is the reason why we do not use their example as it is. Their argument, however, applies unchanged to our example, which satisfies the maintained Assumption 1.

[$C$]. Certainly ($ii$) is not a very convincing argument if player 1 has any reason to think that player 2 will not choose [$C$]. Also, ($i$) implies a very pessimistic point of view; to be sure, $M$ yields at least [0], but it also yields at most [$\varepsilon$]." (ibid., p.62). If 1 had any uncertainty that 2 might be playing $L$ in this game, it would be unreasonable to assume he would not play $U$ for sufficiently small $\varepsilon$. But then it might be unreasonable to rule out $R$, and hence $D$, and ultimately $L$, reinforcing the rationale for $U$. "[...] So it goes, for nothing prevents us from continuing this sort of 'I-think-that-he-thinks-that-I-think-that-he-thinks...' reasoning to the point where all strategy choices appear to be equally reasonable" (ibid., p.62). □

Hence, the strategic uncertainty associated with RCBR (which is represented by the fact that all actions are rationalizable in Example 8) clashes with the sharp message of the maxmin criterion.[15] On the other hand, the latter is grounded on a simple, if extreme, decision theoretic principle. One classical argument to reconcile the two views is to note that the maxmin action ensures optimality (in the expected utility sense) in the eventuality that one's action is leaked to the opponent. Interestingly, the logic behind our result is reminiscent precisely of that argument. In fact, one could think of our model of extensive-form uncertainty precisely as one way of making explicit a 'fear of leaks' that is behind the classical informal intuition for the maxmin criterion.

We point out, however, that whereas the standard 'fear of leaks' argument can be thought of as a first-order beliefs effect, the result in Proposition 2 holds for beliefs effects of any order: an argument similar to that in Section 1.1 can be used to show that, in a zero-sum game such as those in Examples 7 and 8, the maxmin action is the only prediction of RCBR even for types that attach probability *zero* to their action being leaked, and which share arbitrarily many (but finite) orders of mutual belief that there are no leaks.

### 5.3 Mixed Extensions and Robust Refinements

The key aspects of strictly competitive environments are most striking in games that have no pure equilibria, such as the Matching Pennies game (on the left):

| | $H_2$ | $T_2$ |
|---|---|---|
| $H_1$ | $1\ \ -1$ | $-1\ \ 1$ |
| $T_1$ | $-1\ \ 1$ | $1\ \ -1$ |

| | $H_2$ | $M_2$ | $T_2$ |
|---|---|---|---|
| $H_1$ | $1\ \ -1$ | $0\ \ 0$ | $-1\ \ 1$ |
| $M_1$ | $0\ \ 0$ | $0\ \ 0$ | $0\ \ 0$ |
| $T_1$ | $-1\ \ 1$ | $0\ \ 0$ | $1\ \ -1$ |

The tension between RCBR and the maxmin criterion is particularly evident in this case, because the latter yields a unique mixture as its solution, but of course the maxmin

[15] Luce and Raiffa (1957, p.65) connect this tension between the maxmin approach and RCBR to a classical dychotomy in military strategy, with doctrines based on enemy capabilities and intentions, respectively: "A military commander may approach decisions with either of two philosophies. He may select his course of action on the basis of his estimate of what his enem is *able to do* to oppose him. Or, he may make his selection on the basis of his estimate of what his enemy *is going to do*. The former is a doctrine of decision based on enemy capability; the latter, on enemy intentions." (Haywood (1954), pp.365-366).

mixture is optimal for an expected utility agent if and only if he is indifferent over the pure actions in its support (e.g., Pearce (1984), Lemma 1). Hence, the informal intuition that the maxmin mixture might be 'strictly preferred' in this setting is inconsistent with the standard notion of rationality in game theory (namely, expected utility maximization). However, classical arguments in support of playing mixed strategies insist on the idea that they allow to preserve *secrecy* (e.g., von Neumann and Morgenstern (1947, Ch.14 and 17), Luce and Raiffa (1957, Ch.4), Schelling (1960, Ch.7), etc.). Once again, these arguments are reconciled with standard rationality by appealing to an unmodelled 'fear of leaks': choosing the maxmin mixture, rather than the pure actions in its support, ensures secrecy – and hence increases the 'security level' – even if one's strategy is 'found out' by the enemy (of course, this argument presumes that the chosen mixture may be leaked, but not its realization).

These ideas can be explored in our model by adding to the baseline game actions that are equivalent to the maxmin mixtures (as in the game above, on the right).[16] In such a 'mixed extension' of the Matching Pennies game, $(M_1, M_2)$ is the only equilibrium, and the maxmin outcome, but clearly everything is rationalizable. This game, however, does not fall directly under the statement of Proposition 2 because it violates Assumption 1, which rules out $i$'s indifference over his own action when the opponent plays $M_j$. However, it can be shown that our results extend to games that satisfy a weaker version of Assumption 1, which can accommodate mixed extensions of zero-sum games with no pure equilibria, such as the game above. In that extension, the maxmin profile is still the only robust prediction of RCBR.[17]

Mixed extensions may also be considered for coordination games. For instance, the following game consists of a modified version of the Battle of the Sexes, in which players have one extra action that is equivalent to the mixed equilibrium mixture:

| | $B_2$ | $M_2$ | $S_2$ |
|---|---|---|---|
| $B_1$ | 3 1 | 3/4 3/4 | 0 0 |
| $M_1$ | 3/4 3/4 | 3/4 3/4 | 3/4 3/4 |
| $S_1$ | 0 0 | 3/4 3/4 | 1 3 |

This game again violates Assumption 1, and hence it does not fall directly under the statements of Theorem 1. But it can be shown that the robust predictions of RCBR in this case rule out the mixed equilibrium actions $M_i$.[18]

[16]If one player can commit to an objective mixture, to the point (as the argument goes) that its choice might be leaked but not its realization, then it means that an irreversible choice of an objective randomizing device is available to the player. It seems therefore consistent to model the randomizing decise as a pure strategy.

[17]The weaker version of Assumption 1 is the following (see Penta and Zuazo-Garin (2017)): (A.1') For each $i$, for each $A_i' \subseteq A_i$, $\exists! a_i \in A_i' : a_i \in \text{argmax}_{a_i' \in A_i'} u_i(s_j(a_i'); a_i')$ for any $s_j(\cdot)$ s.t. $s_j(a_i) \in \arg\max_{a_j \in A_j} u_j(a_i, a_j)$ for all $a_i$. We don't consider this here because Assumption 1 significantly simplifies the exposition, at the cost of a minor loss in generality.

[18]Interestingly, Kalai and Samet's (1984) 'persistent equilibria' also refine away the mixed equilibrium in this game (see also Myerson (1991)). Arguments for refining away the mixed equilibrium of the Battle of the Sexes had been provided, among others, by Schelling (1960, footnote 18, p.286).

# 6 'Eductive' Coordination via Extensive Form Uncertainty

Understanding the mechanisms by which individuals achieve coordination of behavior and expectations is one of the long-lasting questions in game theory. When individuals interact repeatedly over time, learning theories or evolutionary arguments have been provided to sustain coordination (see, e.g., Fudenberg and Levine (1998), Samuelson (1998) and references therein). But when interactions are one-shot or isolated, or when players have no information about past interaction, their choices can only be guided by their individual reasoning, and whether equilibrium coordination can be achieved is far from understood.

That a purely *eductive* approach, based only on internal inferences, may result in equilibrium coordination is generally met with skepticism. As a result, two main reactions can be found in the literature. At one extreme, non-equilibrium approaches such as rationalizability (e.g., Bernheim (1984) and Pearce (1984)) or level-$k$ theories (e.g., Nagel (1995), Stahl and Wilson (1994, 1995)) have been developed to analyze initial responses in games.[19] At the opposite extreme, other approaches have developed Schelling's (1960) idea of *focal points* (e.g., Sugden (1995), Iriberri and Crawford (2007), etc.), which maintains the equilibrium assumption and shifts the discussion on the mechanisms that bring about coordination to external, non mathematical properties of the game.[20]

In this section we show that higher order uncertainty over the extensive form provides a purely eductive mechanism for equilibrium coordination, based on standard assumptions of RCBR, without appealing to any external theory of focal points.

## 6.1 Generic Coordination

An important class of games in which RCBR ensures equilibrium coordination generically on the universal type space is provided by *Nash commitment games with a unique Pareto efficient equilibrium*. First note that, if $G^*$ satisfies these properties, then both dynamic versions of the game yield the same backward induction outcome (that is, $a^1 = a^2$). Conversely, it is easy to show that if $a^1 = a^2 \equiv a^*$, then $G^*$ is a Nash commitment game, and $a^*$ is its unique efficient equilibrium.

Note that coordination games with Pareto ranked equilibria (sometimes referred to as 'pure' coordination games) and games in which players have the same preferences over strategy profiles (a strong notion of 'common interest' games that is some times adopted), are special cases of games in which $a^1 = a^2$, and hence of Nash commitment games with a unique efficient equilibrium. But, as shown by part ($i$) of Proposition 2, also zero-sum games with a pure equilibrium have this property. The next result therefore applies to all such games:

**Proposition 3 (Generic Coordination-I)** *In Nash commitment games with a unique Pareto efficient equilibrium (equivalently, if $a^1 = a^2 \equiv a^*$), there exists an open and dense subset $T' \subseteq T^*$ such that $R(t) = \{a^*\}$ for all $t \in T'$.*

[19] On level-$k$, see also Costa-Gomes, Crawford and Broseta (2001), Costa-Gomes and Crawford (2006), Crawford and Iriberri (2007), and more recently Alaoui and Penta (2016).

[20] In the words of Schelling (1960, p.108): "it is not being argued that players *do* respond to the non-mathematical properties of the game but that they *ought* to take them into account [...]."

Note that, unlike for Proposition 1, the statement in this proposition refers to the action *profile*: in Proposition 1, individual players choose equilibrium actions, but their choices jointly need not form an equilibrium. Proposition 3, in contrast, ensures that RCBR generically yields an equilibrium *outcome*. In this sense, higher order uncertainty on the extensive form provides a channel through which equilibrium coordination is justified from a purely eductive viewpoint.

Eductive coordination aside, this result is also interesting from the viewpoint of equilibrium refinements. In particular, in *common interest games* with a unique efficient equilibrium (i.e. games such which also satisfy $u_1^*(a) = u_2^*(a)$ for all $a \in A$ – a special case of the class considered in Proposition 3), efficient coordination is a particularly intuitive prediction. Yet, supporting it without involving refinements directly based on efficiency has required in the past surprisingly complex arguments. For instance, Aumann and Sorin (1989) support the efficient equilibrium in this special class of games as the only equilibrium outcome of a repeated game in which one player is uncertain about his opponents' type, and types may have bounded memory. For the same class of games, Lagunoff and Matsui (1997) support the efficient outcome considering a repeated game setting with perfect monitoring in which players choose simultanesouly in the first period, and they alternate after that; Calcagno et al. (2014) instead support the same refinement introducing a pre-play phase in which players prepare the actions that will be implemented at a predetermined deadline, and randomly receive opportunities to revise their actions before the deadline.

In all of these papers, however, the efficient coordination result relies crucially on the fact that earlier moves are observable (albeit they may be revised in the environments of Lagunoff and Matsui (1997) and Calcagno et al. (2014)), and this is common knowledge. In contrast, the efficient coordination in Proposition 3 holds for a generic (open and dense) subset of the type space, regardless of whether players' actions are actually observable.

The fact that the result also holds for a larger class of games suggests that the novel notion of Nash commitment game may provide a more primitive property underlying the possibility of efficient coordination than the restrictive notion of pure common interest.

## 6.2 The Scope of Coordination

Note that, in Example 5, $RP(t) = \{(M, C)\}$ for all types $t \in T^*$ whose first order beliefs attach probability one to the game being static (i.e., for all $t$ such that $\hat{\pi}_{i,1}(t_i)\left[\omega^0\right] = 1$). Hence, whenever players are certain that the game is static, but they face higher order uncertainty on the extensive form, common belief of rationality generically yields $(M, C)$ as the unique solution in that game. So, while in this case equilibrium coordination is not 'generic' in $T^*$ – for instance, because types who are certain of $\omega^1$ would induce outcome $a^1$ – there is still a sense in which equilibrium coordination emerges for a 'large' subset of the universal type space: namely, for all types that share the same first-order beliefs as $t^{CB}\left(\omega^0\right)$.

The next result generalizes this insight to all games in which the robust predictions are unique for the common belief type $t^{CB}\left(\omega^0\right)$, and relates the 'reach' of the coordination result to the number of iterations of the $RP_i^k$ procedure necessary to reach uniqueness. To this end, for any $t \in T^*$ and $k \in \mathbb{N}$, we define the set

$$\Lambda^k(t) := \left\{t' \in T^* : \hat{\pi}_k\left(t'\right) = \hat{\pi}_k(t)\right\}.$$

of all types that share the same $k$-lowest order beliefs as $t$. Then, the following holds:

**Proposition 4 (The Scope of Coordination)** *For any action profile $\hat{a} \in A$ such that $RP^k(t^{CB}(\omega^0)) = RP(t^{CB}(\omega^0)) = \{\hat{a}\}$ for some $k$, we have: (i) $\hat{a} \in NE^*$, and (ii) there is an open set $T' \subseteq T^*$ such that $R(t) = \{\hat{a}\}$ for all $t \in T'$, and $T' \cap \Lambda^k(t^{CB}(\omega^0))$ is dense in $\Lambda^k(t^{CB}(\omega^0))$.*

Hence, 'eductive' coordination occurs for a generic subset of $\Lambda^k\left(t^{CB}\left(\omega^0\right)\right)$, whenever the robust predictions are unique in the static benchmark (i.e., if $RP\left(t^{CB}\left(\omega^0\right)\right) = \{\hat{a}\}$). Moreover, since $\Lambda^k(t) \subseteq \Lambda^{k'}(t)$ for all $t$ and $k > k' \geq 1$, it follows that the 'reach' of coordination is larger if the unique robust prediction $\hat{a}$ is obtained in a smaller number of rounds, and maximally so if this occurs in one round, as in Example 5.

# 7 Asynchronous Moves and Asymmetric Perturbations

If moves are chosen at different points in time, with a commonly known order, then it is plausible to consider asymmetric extensive-form uncertainty: in these settings, it would never be the case that the earlier mover chooses after having observed the action of the later mover. In this section we consider this problem. As we discussed in the introduction, our results provide a rational explanation to the idea that timing alone may determine the attribution of the first-mover advantage, independent of actions' observability.

Let $e \in \{1,2\}$ be the player who moves *earlier*, and $\ell \neq e$ the *later* mover. To represent common knowledge that $\ell$'s action would not be observed by $e$ before making his choice, we consider the smaller space of uncertainty $\Omega^e = \left\{\omega^0, \omega^e\right\}$, and let $T^e$ denote the universal type space generated by $\Omega^e$ – defined in a way analogous to $T^*$ for the larger space of uncertainty $\Omega$ considered in the previous sections. Similarly, both solution concepts $R_i$ and $RP_i$ defined above can be regarded as correspondences defined over $T^e$ in this case.

To identify the robust predictions in $T_i^e$, for each $i$ we define the subset of actions $\mathcal{A}_i := \bigcup_{k\geq 1} \mathcal{A}_i^k$, where $\mathcal{A}_i^1 := \{a_i^e\}$, and for each $k \geq 1$,

$$\mathcal{A}_i^{k+1} := \mathcal{A}_i^k \cup \left\{ a_i \in A_i : \begin{array}{c} \exists \mu^i \in \Delta(\mathcal{A}_j^k), \exists p \in [0,1] \text{ such that:} \\ \{a_i\} = \arg\max_{a_i' \in A_i} \left( p \cdot \sum_{a_j \in A_j} \mu^i[a_j] \cdot u_i^*(a_i', a_j) + (1-p) \cdot u_i^*(a_i', a_j^*(a_i')) \right) \end{array} \right\}.$$

Note that $\mathcal{A}_i$ is basically the same as the set $\mathcal{B}_i$ defined in Section 2, except that we only take profile $a^e$ as the 'seed' (not $a^\ell$). For each $i$, we define the *robust predictions under asymmetric perturbations*, $RP_i^{\mathcal{A}}(\cdot)$, which is obtained replacing the sets $\mathcal{B}_i$ with $\mathcal{A}_i$ in the definition of $RP_i(t_i)$, for each $t_i \in T_i^e$. The next result, analogous to Theorem 1, shows that $RP_i^{\mathcal{A}}$ characterizes the robust predictions in $T_i^e$:

**Theorem 2 (Asymmetric Perturbations)** *For any player $i$, $RP_i^{\mathcal{A}}(\cdot)$ is non-empty valued and upper hemicontinuous on $T_i^e$. Moreover, for any finite type $t_i \in T_i^e$ and any strategy $s_i \in RP_i^{\mathcal{A}}(t_i)$, there exists a sequence of finite types $(t_i^\nu)_{\nu \in \mathbb{N}} \subseteq T_i^e$ with limit $t_i$ and such that $R_i(t_i^\nu) = RP_i^{\mathcal{A}}(t_i^\nu) = \{s_i\}$ for all $\nu \in \mathbb{N}$.*

The following corollary states properties of $RP_i^{\mathcal{A}}$ analogous to those of Corollaries 1-3:

**Corollary 5** *For any player $i$, the following holds: (i) $R_i$ is* unique *and coincides with $RP_i^{\mathcal{A}}$ over an open and dense set of types $T_i' \subseteq T_i^e$. (ii) No proper refinement of $RP_i^{\mathcal{A}}$ is upper hemicontinuous on $T_i^e$. (iii) For any $t_i$, if there exists a sequence $(t_i^\nu)_{\nu \in \mathbb{N}} \subseteq T_i^e$ with limit $t_i$ such that $R_i(t_i^\nu) = \{s_i\}$ for all $\nu \in \mathbb{N}$, then $s_i \in RP_i^{\mathcal{A}}(t_i)$*

Hence, in the universal type space $T^e$ generated by the asymmetric space of uncertainty $\Omega^e \subseteq \Omega$, the refinement $RP_i^{\mathcal{A}}(\cdot) \subseteq R_i(\cdot) \cap \mathcal{A}_i$ plays the same role as $RP_i$ in the universal type space $T^*$ we considered in the previous sections. In the rest of this section we explore special implications of the results in Theorem 2.

## 7.1 Pervasiveness of the First-Mover Advantage

Theorem 2 has especially interesting implications in Nash commitment games (see Def. 3):

**Proposition 5 (Pervasiveness of First-Mover Advantage)** *If $G^*$ is a Nash commitment game, there is an open and dense subset of types $T_i' \subseteq T_i^e$ such that, for all $t_i \in T_i'$, $R_i(t_i) = \{a_i^e\}$ if $\hat{\theta}_i(t_i) = \theta_i'$, and $R_i(t_i) = \{a^*(\cdot)\}$ if $\hat{\theta}_i(t_i) = \theta_i''$.*

That is, whenever higher order beliefs do not rule out the possibility that player $e$ has a first-mover advantage, then (generically) $e$ has a *de facto* first-mover advantage. In this sense, we say that a first-mover advantage is *pervasive.*

This result has important strategic implications, in that it implies that in the presence of higher order uncertainty about the extensive form (i.e., generically on the universal type space), the *timing* of moves alone may determine the attribution of strategic advantage. This is in sharp contrast with the received game theoretic intuition, which is that commitment and observability are key, not timing per se. Higher order uncertainty overturns the received intuition, because a commonly known timing of moves induces a *one-sided uncertainty* about the extensive form. This, generically, induces the same outcome as if the game had perfect information, even if the action is *not* observed, and not believed to be observed for arbitrarily many (but finite) orders.

The notion that timing has strategic importance, beyond commitment and observability, has been discussed by Kreps (1990), and the idea has received strong support by the experimental literature (see, e.g., Camerer (2003) and references therein). Cooper et al. (1993), in particular, study a Battle of the Sexes with asynchronous moves, in which players are told that the second mover does not observe the first mover's action, so that the strategic form is identical to a static game in which moves are played simultaneously. Yet, Cooper et al. (1993) show that, unlike in the simultaneous move treatment, equilibrium coordination is typically achieved (on the backward induction outcome) in the asynchronous treatment, as if the game had perfect information. This is in line with the *Kreps hypothesis* (Kreps (1990), pp.100-101), but clearly at odds with the received game theoretic wisdom. To the best of our knowledge Proposition 5 is the first result which can make sense of this solid experimental evidence, without appealing to behavioral theories or notions of bounded rationality, while maintaining non-observability of the earlier mover's action.[21]

This is not to say that the logic of our results – which involves possibly very complex higher order reasoning – necessarily provides a behaviorally accurate model of strategic thinking (cf., Crawford, Costa-Gomes and Iriberri (2013)). But our results have methodological implications for the behavioral and experimental literature nonetheless. First, because they show that standard (i.e., non-behavioral) notions such as higher order uncertainty and RCBR may explain patterns of behavior that are seemingly at odds with classical game theoretic reasoning. Hence, once the assumption of common knowledge of the extensive form is perturbed, standard game theoretic analysis may provide an effective *as if* model of how timing impacts players' strategic reasoning. Second, because our results suggest that – contrary to the received wisdom – systematic biases may be associated with aspects of experimental design, such as the timing of moves, which have an asymmetric impact on higher order uncertainty. Finally, while Theorem 2 in general relies on the entire belief hierarchy, Proposition 7 below shows that the robust predictions are often pinned down by lower order beliefs.

We conclude this discussion by noting that the result of Proposition 5 in fact applies to any situation that generates a situation of asymmetric uncertainty. While timing is a natural source of asymmetry, it need not be the only one. For instance, in the context of a duopoly, it may be common knowledge that one firm has succesfully committed to ignoring the other's choice, or that one firm is succesful at preventing leaks, but there may be higher order uncertainty concerning the other firm. The results in this section would also apply to these cases, letting these firms take the roles of players $e$ and $\ell$, respectively.[22]

### 7.2 Robustness of the First-Mover Advantage

The message of Proposition 5, that the first-mover advantage is pervasive, may appear to be in sharp contrast with Bagwell (1995), who argued instead that the first-mover advantage is rather fragile.[23] There are two main differences that explain the discrepancy between our results and Bagwell's (1995). The most important difference is that Bagwell (1995)

[21]Amershi, Sadanand and Sadanand (1989, 1992) developed solution concepts that assign a specific role to timing as a coordinating device, and hence they appeal to 'external' considerations. Their approach is perhaps more in line with what Kreps (1990) seemed to suggest: to look for an answer in a theory of focal points. In contrast, the role of timing in our results stems precisely as *the* (generic) implication of RCBR, when there is higher order uncertainty on the observability of the first mover's action, with no need to involve focal points or any external consideration. As already mentioned, the result does not require that actions are actually observable. As discussed in Footnote 5, this is an important difference with the respect to the works by Lagunoff and Matusi (1997) and Ambrus and Ishii (2015).

[22]We are thankful to Glenn Ellison for this observation.

[23]This interpretation of Bagwell's (1995) result has been criticized, among others, by Van Damme and Hurkens (1997), who showed that the perturbed model in Bagwell (1995) admits a mixed equilibrium which converges to the the backward induction solution as the perturbations vanish. Hence, the apparent fragility of the first-mover advantage in Bagwell (1995) stems from a particular equilibrium selection in the perturbed model. For an earlier analysis of noisy information on earlier moves, see Bonanno (1992).

considers perturbations of the benchmark with perfect information that are such that the second mover observes the first mover's action with noise. In contrast, in our perturbations, the second mover either observes the action without noise, or it doesn't observe it at all. This information structure could be cast in Bagwell's (1995) model, but allowing for non-overlapping distributions of signals conditional on the different actions, which are ruled out by his common support assumption. At the same time, extensive-form uncertainty in our case may enter through higher order beliefs alone, whereas Bagwell's (1995) perturbations involve the first order beliefs as well. The second difference, although less important, is that Bagwell's (1995) results refer to non-Nash commitment games. The reason why this difference is somewhat secondary is that, in a non-Nash commitment game, the Stackelberg action is uniquely rationalizable for the common belief type $t_e^{CB}(\omega^e)$ in our model, and in fact it is also locally robust. So, for non-Nash commitment games, the first-mover advantage may not be 'pervasive', but it would still be robust around $t^{CB}(\omega^e)$. Thus, the real source of the discrepancy is the assumption on the signals structure discussed above. The next proposition, which in fact follows immediately from Theorem 2, formalizes the idea:

**Proposition 6 (Robustness of First-Mover Advantage)** *For any $G^*$, there is an open neighborhood of $t^{CB}(\omega^e)$ in which the only rationalizable solution induces the backward induction outcome of the game with perfect information in which $e$ moves first.*

### 7.3 Eductive Coordination via Asynchronous Moves

We first note that the result in Proposition 5 implies that, in environments with asynchronous moves, higher order uncertainty about the extensive form generically yields equilibrium coordination, even in games in which players do not share the same ranking of the equilibrium outcomes, and hence there is no single Pareto efficient equilibrium. Thus, when moves are asynchronous or more generally under asymmetric perturbations, the eductive coordination result of Proposition 3 extends to all Nash commitment games.

We conclude this section with a result that characterizes the scope of coordination in general games with asymmetric moves, similar to what we did in Proposition 4, but for the case of asymmetric perturbations. To this end, for any $t \in T^e$ and for any $k \in \mathbb{N}$, let

$$\Lambda^{e,k}(t) := \left\{t' \in T^e : \hat{\pi}_k\left(t'\right) = \hat{\pi}_k(t)\right\}.$$

**Proposition 7 (Eductive Coordination with Asynchronous Moves)** *Let $t^{CB}\left(\omega^0\right) \in T^e$ denote the type profile representing common belief in $\omega^0$. Then, for any action profile $\hat{a} \in A$ such that $RP^{\mathcal{A},k}\left(t^{CB}\left(\omega^0\right)\right) = RP\left(t^{CB}\left(\omega^0\right)\right) = \{\hat{a}\}$ for some $k$, we have: (i) $\hat{a} \in NE^*$, and (ii) there is an open set $T' \subseteq T^e$ such that $R(t) = \{\hat{a}\}$ for all $t \in T'$ and $T' \cap \Lambda^{e,k}\left(t^{CB}\left(\omega^0\right)\right)$ is dense in $\Lambda^{e,k}\left(t^{CB}\left(\omega^0\right)\right)$.*

Thus, similar to Proposition 4, the reach of the eductive coordination result via extensive-form uncertainty is larger, the lower the number of iterations of the $RP^{\mathcal{A}}$ operator needed to attain uniqueness. Hence, when uniqueness is reached in $k$ steps, eductive coordination is attained by only relying on the lower $k$ order beliefs.

## 8 Concluding Remarks

As we discussed in Section 1.2, extensive-form uncertainty and robustness are relatively unexplored questions in game theory, which could be approached from many angles. In this paper we studied the implications of perturbing common knowledge assumptions on the observability of players' actions in two-player games. In contrast with the famous results by Weinstein and Yildiz (2007), who studied an analogous question for payoff uncertainty in static games, we found that the 'robust predictions' in our sense generally support a robust refinement of rationalizability, with interesting implications in important classes of games: (*i*) in 'agreement games with constant disagreement payoffs' (a generalization of Harsanyi (11981) and Kalai and Samet's (1984) 'unanimity games'), the robust predictions select one of the backward induction actions generically on the universal type space, and hence they rule out inefficient equilibrium actions; (*ii*) in strictly competitive games, the robust predictions uniquely select the maxmin solution, thereby resolving a classical tension between RCBR and the maxmin logic which has long been discussed (e.g., Luce and Raiffa (1957); Schelling (1960); (*iii*) in 'Nash commitment games with a unique efficient equilibrium' (which include as special cases both zero-sum games with a pure equilibrium and Aumann and Sorin's (1989) common interest games), the robust predictions ensure efficient coordination of behavior, thereby showing that higher order uncertainty over the extensive form may serve as a mechanism for *eductive coordination* (Binmore, 1987-88; see also Guesnerie, 2005).

We also applied our approach to situations with asynchronous moves, in which the order of moves is common knowledge, but not their non-observability. We found that, in any 'Nash commitment game', the robust predictions uniquely select the equilibrium outcome which is most favorable to the first mover generically on the universal type space. Hence, a first-mover advantage arises in these games whenever higher order beliefs do not rule out it exists: no matter how close we are to having common knowledge that actions are *not* observable, the unique outcome consitent with rationality and common belief in rationality is the same *as if* we had common knowledge that the earlier mover's action is observable. This suggests that, in the presence of higher order uncertainty on the observability of actions, *timing* of moves alone may determine the attribution of the strategic advantage – a message which is clearly at odds with the received game theoretic intuition, which therefore relies crucially on the maintained assumptions of common knowledge of the extensive form. These results also provide a rational basis for the *Kreps hypothesis* (Kreps, 1990), which maintains that timing may have strategic importance beyond commitment and observability – an idea which has found extensive experimental support (see Camerer (2003) and references therein), but which has been difficult to reconcile with standard game theoretic analysis.

The breadth of results allowed by our robustness exercise suggests that further exploring the problem of extensive-form uncertainty may prove to be a fertile direction for future research. From a more applied perspective, it would be interesting to further explore the implications of our analysis to other classes of games, not covered by our results in Sections 5-7. From a more theoretical viewpoint, it would be important to extend our analysis beyond two-player games. The most obvious challenge in doing this is represented by the fact that, as the number of players increase, the number of dynamic games that could be obtained by different permutations of the players grows very quickly. It would also be interesting to identify the robust predictions when actions not only may be observed perfectly or not at all, but also allowing partial observability. It seems clear that, in that case, the 'robust predictions' would be less sharp than the ones we obtained in general games, but an exact characterization is far from obvious.

More broadly, different notions of extensive-form robustness can be developed, mimicking the several notions of robustness which have been developed by the literature on payoff uncertainty. For instance, Ely and Doval (2016) and Makris and Renou (2018) study environments in which an external analyst has limited information on the extensive form, and seek to characterize all the equilibirum outcomes which could be generated in all dynamic games consistent with the analyst's information (the actual extensive form, however, is common knowledge among players). Similarly, Peters (2015) characterizes the distributions over outcomes an analyst can expect when he only knows that the actual extensive form belongs to a certain (large) class. In this sense, the exercises in these papers are analogous to those pursued, in environments with payoff uncertainty, by Bergemann and Morris (2013, 2016) in game theory and by Bergemann and Morris (2005, 2009) and Penta (2015) in mechanism design. In contrast, the notion of extensive-form robustness we considered in this paper is 'local' in the sense that we only look at the impact of small perturbations of players' belief hierarchies over the extensive form. From that viewpoint it is similar to the approach of Weinstein and Yildiz (2007) in game theory, and of Oury and Tercieux (2012) in mechanism design.[24] Intermediate notions of robustness with payoff uncertainty, which have been put forward in the mechanism design literature (e.g., Artemov, Kunimoto and Serrano (2013), Ollar and Penta (2017)), may suggest further directions of research on extensive-form robustness.

In conclusion, the problem of extensive-form robustness is very broad. Future work may study different classes of games, richer spaces of uncertainty, different notions of robustness (local, global, intermediate as well as alternative topologies), an so on. We provided one of the first attempts at understanding this question, but the modeling possibilities are very rich, and suggest many promising directions for future research.

# Appendix

## A Universal Type Space

### A.1 General Case

To complete the description of the strategic situation, players' beliefs about what they do not know must be specified. That is, for every $i$, his beliefs about $\Omega$ (*first order beliefs*), his beliefs about $\Omega$ and the opponent's first order beliefs (*second order beliefs*), and so on. Remember that each $\theta_i \in \Theta_i$ is a subset of $\Omega$. Then, the set of possible *first order beliefs*

[24] Kaji and Morris (1997) can also be thought of as a 'local' robustness exercise for environment with payoff uncertainty, but in a different topology than the papers discussed above.

*consistent with* $\theta_i$ is defined as $Z_{i,1}(\theta_i) := \Delta(\theta_i)$. Also define player $j$'s first order beliefs that are consistent with $i$'s information $\theta_i$ as:

$$Z_{j,1}(\theta_i) := \{\pi_{j,1} \in Z_{j,1}(\theta_j) : \theta_j \cap \theta_i \neq \emptyset\}.$$

These are the first order beliefs of player $j$ that are not inconsistent with player $i$'s information $\theta_i$; and thus, the only ones that might eventually receive positive probability by a belief consistent with $\theta_i$. Recursively, also define for any $k \in \mathbb{N}$,

$$Z_{i,k+1}(\theta_i) := \left\{(\pi_{i,n})_{n=1}^{k+1} \in Z_{i,k}(\theta_i) \times \Delta(\theta_i \times Z_{j,k}(\theta_i)) : \operatorname{marg}_{\theta_i \times Z_{j,k-1}(\theta_i)} \pi_{i,k+1} = \pi_{i,k}\right\},$$
$$Z_{j,k+1}(\theta_i) := \left\{(\pi_{j,n})_{n=1}^{k+1} \in Z_{j,k+1}(\theta_j) : \theta_j \cap \theta_i \neq \emptyset\right\}.$$

Thus, type $\theta_i$'s *first order beliefs* are elements of $\Delta(\theta_i)$. An element of $\Delta(\theta_i \times Z_j^{k-1}(\theta_i))$ is a type $\theta_i$'s *k-th order belief*. The set of (*collectively coherent*) *belief hierarchies* for type $\theta_i$ is then defined as:

$$H_i(\theta_i) := \left\{\pi_i \in Z_{i,1}(\theta_i) \times \prod_{k \in \mathbb{N}} \Delta(\theta_i \times Z_{j,k}(\theta_i)) : \forall k \in \mathbb{N}, (\pi_{i,n})_{n=1}^{k} \in Z_{i,k}(\theta_i)\right\},$$

and the set of all (*consistent*) *information-hierarchy* pairs, as,

$$T_i^* := \bigcup_{\theta_i \in \Theta_i} \{\theta_i\} \times H_i(\theta_i).$$

It follows from Mertens and Zamir (1985) that when $T_i^*$ is endowed with the product topology there exists a homeomorphism $\tau_i^* : T_i^* \longrightarrow \Delta(T_j^*)$ that preserves beliefs of all orders; i.e., such that for every information-hierarchy pair $(\theta_i, \pi_i)$ we have both that:

(*i*) $\pi_{i,1}[\omega] = \tau_i^*(t_i)[\{(\theta_j, \pi_j) \in T_j^* | \theta_i \cap \theta_j = \{\omega\}\}]$ for any state $\omega$.

(*ii*) $\pi_{i,k}[E] = \tau_i^*(t_i)[\mathrm{Proj}^{-1}_{\theta_i \times Z_{i,k-1}}(E)]$ for any measurable $E \subseteq \theta_i \times Z_{i,k-1}(\theta_i)$ and any $k > 1$.

Hence, the tuple $\mathcal{T}^* := (T_i^*, \hat{\theta}_i, \tau_i^*)_{i \in I}$, where $\hat{\theta}_i(\theta_i, \pi_i) := \theta_i$ for every information-hierarchy pair $(\theta_i, \pi_i)$, is an information-based type space. It will be referred to as the (*information-based*) *universal type space.*

Now, every type $t_i$ from a type space $\mathcal{T} = (T_i, \hat{\theta}_i, \tau_i)_{i \in I}$, induces an consistent information-hierarchy pair determined by information $\hat{\theta}_i(t_i)$ and:

- First order beliefs specified by map $\hat{\pi}_{i,1} : T_i \to \Delta(\Omega)$, where for any $E \subseteq \Omega$,

$$\hat{\pi}_{i,1}(t_i)[E] := \tau_i(t_i)\left[\left\{t_j \in T_j : \hat{\theta}_i(t_i) \cap \hat{\theta}_j(t_j) \subseteq E\right\}\right],$$

- Higher order beliefs specified by, for each $k > 1$, map $\hat{\pi}_{i,k} : T_i \to \Delta(\Omega \times Z_{j,k-1})$, where for any measurable $E \subseteq \Omega \times Z_{j,k-1}$,

$$\hat{\pi}_{i,k}(t_i)[E] := \tau_i(t_i)\left[\left\{t_j \in T_j : \hat{\theta}_i(t_i) \cap \hat{\theta}_j(t_j) \times \{\hat{\pi}_{j,k-1}(t_j)\} \subseteq E\right\}\right].$$

Then, continuous map $\phi_i : T_i \to T_i^*$ given by $t_i \mapsto (\hat{\theta}_i(t_i), \hat{\pi}_i(t_i))$ where $\hat{\pi}_i(t_i) := (\hat{\pi}_{i,n}(t_i))_{n\in\mathbb{N}}$ assigns to each type in an information-based type space its corresponding information-hierarchy pair. Mertens and Zamir (1985) showed that for arbitrary non-redundant information-based type space $\mathcal{T} = (T_i, \hat{\theta}_i, \tau_i)_{i\in I}$,[25] set $\phi_i(T_i)$ is a *belief-closed* subset of $T_i^*$, in the sense that for every type $t_i \in \phi_i(T_i)$ belief $\tau_i^*(t_i)$ assigns full probability to $\phi_j(T_j)$. A type $t_i \in T_i^*$ is *finite* if it belongs to a finite belief-closed subset of $T_i^*$.

## A.2 Asynchronous Moves

We replicate now the construction above for the case in which it is commonly known that a player does not move first, i.e., for a model in which state $\omega^j$ is excluded both in terms of knowledge and beliefs for some player $j$. To this end, let $e$ denote the player who might move first. The set of possible states is now $\Omega^e = \{\omega^0, \omega^e\}$. The knowledge partitions are,

$$\Theta_i^e := \begin{cases} \{\{\omega^0, \omega^e\}\} & \text{if } i = e, \\ \{\{\omega^0\}, \{\omega^e\}\} & \text{otherwise.} \end{cases}$$

Thus, player $e$ cannot distinguish between the two states, but her opponent can; it follows that every type of player $e$ is a $\theta_i'$-type (has information $\{\omega^0, \omega^e\}$), while types of player $j \neq e$ can be either $\theta_i'$-types or $\theta_i''$-types (depending whether they have information $\{\omega^0\}$ or $\{\omega^e\}$, respectively). Following exactly the same steps as in the previous section we can construct different $k$-th order belief spaces $Z_{i,k}^e(\theta_i)$ for types $\theta_i \in \Theta_i^e$, the set of consistent information-belief hierarchy pairs $T_i^e := \bigcup_{\theta_i \in \Theta_i^e} \{\theta_i\} \times H_i^e(\theta_i)$ and universal type space $\mathcal{T}^e := (T_i^e, \hat{\theta}_i, \tau_i^e)_{i\in I}$. Lemma 5 below shows that these types can be identified with the subset of types in $T^*$ that exhibit common belief in player $j \neq e$ not moving first, denoted by $T_i^*(e)$ and formally defined as $T_i^*(e) := \bigcap_{k\in\mathbb{N}} B_i^k(\Omega^e)$, where:

- $B_i^1(\Omega^e) := \{t_i \in T_i^* : \tau_i^*(t_i)[\Omega^e \times T_j^*] = 1\}$.
- $B_i^k(\Omega^e) := \{t_i \in T_i^* : \tau_i^*(t_i)[B_j^{k-1}(\Omega^e)] = 1\}$ recursively, for any $k \in \mathbb{N}$.

**Lemma 5** *There exists an embedding $\phi_i^e : T_i^e \to T_i^*$ such that $\phi_i^e(T_i^e) = T_i^*(e)$.*

**Proof.** We are going to show that such embedding is, precisely, $\phi_i^e = \phi_i$. We keep the superscript in $\phi_i^e$ to emphasize the particular embedding we are representing. Note first that even though $\Omega^e \subseteq \Omega$, $\mathcal{T}^e$ is not a $(\Omega, \Theta)$-based type space, because $\Theta_i^e \neq \Theta_i$ for both players. To avoid this issue, Define first, for each player $i$ map $\hat{\theta}_i^e : T_i^e \to \Theta_i$ as follows:

$$t_i \mapsto \begin{cases} \hat{\theta}_i(t_i) \cup \{\omega^\ell\} & \text{if } \hat{\theta}_i(t_i) = \{\omega^0\} \\ \hat{\theta}_i(t_i) & \text{otherwise.} \end{cases}$$

Then, $\bar{\mathcal{T}}^e := (T_i^e, \hat{\theta}_i^e, \tau_i^e)_{i\in I}$ is a $(\Omega, \Theta)$-based type space and therefore, we know from Mertens and Zamir (1985) that if $\bar{\mathcal{T}}^e$ is non-redundant, then $T_i^e$ is homeomorphic to $\phi_i^e(T_i^e)$. That $\bar{\mathcal{T}}^e$ is non-redundant follows easily from the facts that $\hat{\theta}_i^e(t_i) = \hat{\theta}_i^e(t_i')$ and $\hat{\pi}_{i,k}(t_i) = \hat{\pi}_{i,k}(t_i')$ for

[25] Type space $\mathcal{T}$ is *non-redundant* if for any player $i$ and any two types $t_i \neq t_i'$, $\phi_i(t_i) \neq \phi_i(t_i')$.

any $k \in \mathbb{N}$ and any two types $t_i = (\theta_i, \pi_i), t'_i = (\theta_i, \pi'_i) \in T_i^e$ such that $\pi_{i,k} = \pi'_{i,k}$. Hence, $T_i^e$ is homeomorphic to $\phi_i^e(T_i^e)$. That $\phi_i^e(T_i^e) \subseteq T_i^*(e)$ follows from the facts that $\phi_i^e(T_i^e) \subseteq B_i^1(\Omega^e)$ and $\phi_i^e(T_i^e) \times \phi_j^e(T_j^e)$ is belief-closed. Finally, $T_i^*(e) \subseteq \phi_i^e(T_i^e)$ is a consequence of:

$$B_i^k(\omega^e) \subseteq \bigcup_{\theta_i \in \Theta_i} \mathrm{Proj}_{Z_{i,k}(\theta_i)}(\phi_i^e(T_i^e)) \times \prod_{n \geq k} \Delta(\theta_i \times Z_{j,n}(\theta_i)).$$

for any $k \in \mathbb{N}$. ■

# B Basic Properties of the Solution Concepts

In the following, we let $R^{\mathcal{T}}$ denote the notion of interim rationalizability defined in the text, for a given type space $\mathcal{T}$. For the universal type space defined above, we drop superscripts and simply write $R$. Then, we have:

**Proof of Lemma 1.** Let's check that for any player $i$, $R_i^{\mathcal{T},k}(t_i) = R_i^k(\phi_i(t_i))$ for any $k \in \mathbb{N}$ and any type $t_i \in T_i$. We proceed by induction on $k$. The initial step ($k = 1$) holds by definition. For the inductive step, suppose that $k > 1$ is such that the claim holds and fix player $i$ and type $t_i \in T_i$. Then, for the left to right inclusion pick strategy $s_i \in R_i^{\mathcal{T},k+1}(t_i)$ and conjecture $\mu_i \in \Delta(R_j^{\mathcal{T},k})$ such that $\mathrm{marg}_{T_j}\mu_i = \tau_i(t_i)$ and $s_i \in BR_i(\mu_i; t_i)$. Define now conjecture $\mu_i^* \in \Delta(R_j^k)$ as follows:[26]

$$\mu_i^*[E] := \mu_i\left[\left\{(t_j, s_j) \in R_j^{\mathcal{T},k} : (\phi_j(t_j), s_j) \in E\right\}\right],$$

for any measurable $E \subseteq R_j^k$. Obviously, $\mu_i^*$ induces $\tau_i^*(\phi_i(t_i))$ by marginalization. In addition, for any type $t_j$ consistent with $t_i$ it holds that $\omega(t_i, t_j) = \omega(\phi_i(t_i), \phi_j(t_j))$; thus, it follows that $BR_i(\mu_i; t_i) = BR_i(\mu_i^*; t_i)$. Jointly, these conditions imply that $s_i \in R_i^{k+1}(\phi_i(t_i))$.

For the right to left inclusion, pick strategy $s_i \in R_i^{k+1}(\phi_i(t_i))$ and conjecture $\mu_i^* \in \Delta(R_j^k)$ such that $\mathrm{marg}_{T_j^*}\mu_i^* = \tau_i^*(\phi_i(t_i))$ and $s_i \in BR_i(\mu_i^*; \phi_i(t_i))$. Now, we know by the Disintegration Theorem that there exist maps $F_i : T_j^* \to \Delta(T_j)$ and $G_i : T_j^* \to \Delta(T_j^* \times S_j)$ such that:[27]

$$\tau_i(t_i) = \int_{t_j \in T_j^*} F_i(t_j)[\,\cdot\,]\mathrm{d}\tau_i^*(\phi_i(t_i)) \text{ and } \mu_i^* = \int_{t_j \in T_j^*} G_i(t_j)[\,\cdot\,]\mathrm{d}\tau_i^*(\phi_i(t_i)),$$

and,

$$F_i(t_j)[T_j \setminus \phi_j^{-1}(t_j)] = G_i(t_j)[T_j^* \setminus \{t_j\} \times S_j] = 0,$$

for $\tau_i^*(\phi_i(t_i))$-almost every $t_j$. Thus, roughly speaking, measures $F_i(t_j)$ and $G_i(t_j)$ can be interpreted as conditional probability measures $\tau_i(t_i)[\,\cdot\,|\phi_j^{-1}(t_j)]$ and $\mu_i^*[\,\cdot\,|t_j]$, respectively.

[26] Continuity of $\phi_j$ ensures that the following is a well-defined probability measure.

[27] See Theorem 5.3.1 in Ambrosio, Gigli and Savaré (2005), p. 121). We are working with compact and metrizable space; thus, in particular, all of them are Polish and hence, Radon spaces.

Now, define conjecture $\mu_i \in \Delta(T_j \times S_j)$ by setting:

$$\mu_i[E] := \int_{t_j^* \in T_j^*} \left( \int_{t_j \in \phi_j^{-1}(\bar{t}_j^*)} G_i(\bar{t}_j^*) \left[ \{t_j^*\} \times \{s_j \in S_j : (t_j, s_j) \in E\} \right] \mathrm{d}F_i(t_j^*) \right) \mathrm{d}\tau_i^*(\phi_i(t_i)),$$

for any measurable $E \subseteq T_j \times S_j$. It is easy to check that $\mu_i$ is a well-defined element of $\Delta(T_j \times S_j)$ that induces $\tau_i(t_i)$ and has the same marginal on $S_j$ than $\mu_i^*$. It follows from the last two properties that $s_i$ is a best reply to $\mu_i$. It also follows from simple algebra and the induction hypothesis that $\mu_i$ puts full probability on $\Delta(R_j^k)$. Thus, we conclude that $s_i \in R_i^{\mathcal{T},k+1}(t_i)$. ■

**Proof of Lemma 2.** We verify that $R_i^k(\cdot)$ is upper-hemicontinuous for any player $i$ and any $k \in \mathbb{N}$. The initial step ($k = 1$) holds trivially.[28] For the inductive step, suppose that $k \geq 1$ is such that the claim holds; we verify next that it also holds for $k+1$. Now, pick convergent sequence $((t_i^\nu, s_i^\nu))_{\nu \in \mathbb{N}} \subseteq R_i^{k+1}$ with limit $(t_i, s_i)$. Then, we know that there exists a sequence $(\mu_i^\nu)_{\nu \in \mathbb{N}} \subseteq \Delta(R_j^k)$ such that, for any $\nu \in \mathbb{N}$: $(i)$ $\mathrm{marg}_{T_j^*} \mu_i^\nu = \tau_i^*(t_i^\nu)$ and $(ii)$ $s_i^\nu \in BR_i\left(\mu_i^\nu; t_i^\nu\right)$. Since, by the inductive hypothesis, $\Delta(R_j^k)$ is compact, we know that there exists a convergent subsequence $(\mu_i^{\nu_m})_{m \in \mathbb{N}} \subseteq (\mu_i^\nu)_{\nu \in \mathbb{N}}$, whose limit we denote by $\mu_i$. Then, since subsequence $(t_i^{\nu_m})_{m \in \mathbb{N}}$ necessarily converges to $t_i$ and maps $\mathrm{marg}_{T_j^*} : \Delta(R_j^k) \to \Delta(T_j^*)$ and $\tau_i^*$ are continuous, we know that $\mu_i$ induces $\tau_i^*(t_i)$ by marginalization. Furthermore, since $(s_i^{\nu_m})_{m \in \mathbb{N}}$ converges to $s_i$ and $BR_i$ is upper-hemicontinuous, it follows from $(ii)$ that $s_i \in BR_i(\mu_i; t_i)$. Hence, $s_i \in R_i^k(t_i)$. ■

**Lemma 6 (Sequential rationality)** *For any player $i$, any $\theta_i''$-type $t_i$ and any strategy $s_i \in R_i^1(t_i)$, it holds that $s_i(a_j) = a_i^*(a_j)$ for all $a_j \in A_j$.*

**Proof.** Fix player $i$ and $\theta_i''$-type $t_i$, and pick strategy $s_i \in R_i^1(t_i)$ and conjecture $\mu_i \in \Delta(R_j^0)$ that satisfies conditions $(i)$–$(iii)$ in the definition of $R_i$. By contradiction, suppose that there exists some $\hat{a}_j \in A_j$ such that $s_i(\hat{a}_j) \neq a_i^*(\hat{a}_j)$. Let $s_i' \in S_i$ be defined as $s_i'(\hat{a}_j) := a_i^*(\hat{a}_j)$, and $s_i'(a_j) := s_i(a_j)$ for all $a_j \neq \hat{a}_j$. Then:

$$\int_{T_j^* \times S_j} u_i(s_i', s_j, \omega(t_i, t_j)) \mathrm{d}\mu_i - \int_{T_j^* \times S_j} u_i(s_i, s_j, \omega(t_i, t_j)) \mathrm{d}\mu_i =$$
$$= \mu_i \left[ \left\{ (t_j, s_j) : \omega\left(t_i, t_j\right) = \omega^j \text{ and } s_j = \hat{a}_j \right\} \right] \cdot \left( u_i^*(a_i^*(\hat{a}_j), \hat{a}_j) - u_i^*(s_i(\hat{a}_j), \hat{a}_j) \right) > 0,$$

and thus, $s_i \notin BR_i(\mu_i; t_i)$, which is a contradiction. ■

**Lemma 7 (Best-Response Property)** *Set $RP(t^{CB}(\omega^0))$ is the largest product subset of $R^* \cap \mathcal{B}$ satisfying the best-response property.*

**Proof.** Since $RP(t^{CB}(\omega^0))$ obviously satisfies the best-response property and is a subset of $R^* \cap \mathcal{B}$, it suffices to check that for any product set $Q \subseteq R^* \cap \mathcal{B}$ satisfying the best-response property, $Q \subseteq RP^k(t^{CB}(\omega^0))$ for any $k \geq 0$. We proceed inductively. The initial

[28] Just notice that $R_i^1(\cdot)$ only depends on player $i$'s information and that, if $\{t_i^\nu\}_{\nu \in \mathbb{N}}$ converges to $t_i$, then, finiteness of $\Theta_i$ implies that there exists some $\bar{\nu} \in \mathbb{N}$ such that $\hat{\theta}_i(t_i^\nu) = \hat{\theta}_i(t_i)$ for any $\nu \geq \bar{\nu}$.

step ($k=0$) holds trivially; for the inductive argument suppose that $k \geq 0$ is such that the claim holds. Then, fix product subset $Q \subseteq R^* \cap \mathcal{B}$ satisfying the best-response property and player $i$, and pick arbitrary $s_i \in Q_i$. Then, since $s_i \in Q_i$, we know that there exists some conjecture $\mu_i \in \Delta(Q_j)$ such that $s_i$ is a best reply to $\mu_i$ in the (normal form) baseline game. Conjecture $\mu_i$ induces element $\bar{\mu}_i \in \Delta(T_j \times S_j)$ in the obvious way: just set $\bar{\mu}_i := 1_{\{t_j^{CB}(\omega^0)\}} \times \mu_i$, where, with some abuse of notation, we consider $\mu_i$ as extended to $\Delta(S_j)$ by letting it assign probability 0 to any strategy $s_j \in S_j(\omega^j)$. Clearly, $\bar{\mu}_i$ induces $t_i^{CB}(\omega^0)$ and satisfies that $s_i \in BR_i(\bar{\mu}_i; t_i^{CB}(\omega^0))$. Note in addition that since, by the induction hypothesis, $Q_j \subseteq RP_j^k(t_j^{CB}(\omega^0)$, then we also have that $\bar{\mu}_i \in \Delta(RP_j^k)$. Thus, it holds that $s_i \in RP_i^{k+1}(t_i^{CB}(\omega^0))$ and therefore, we conclude that $Q \subseteq RP^{k+1}(t^{CB}(\omega^0))$. ■

# C Robust Predictions

We introduce first an auxiliary solution concept, $SP_i$, which mimics the iterative procedure in the definition of $RP_i$ except for the fact that it deletes not only non-best replies, but also the strategies that are not non-strict best replies. Consider for each player $i$:

$$SP_i^0(t_i) := \begin{cases} \mathcal{B}_i & \text{if } \theta_i(t_i) = \theta_i^0, \\ \{a^*(\cdot)\} & \text{otherwise,} \end{cases}$$

and $SP_i^0 := \{(t_i, s_i) \in T_i^* \times S_i : s_i \in SP_i^0(t_i)\}$. For all the subsequent rounds, both kind of types only perform iterated strict dominance; thus, inductively, for all $k \geq 0$, having defined $SP_j^k := \{(t_j, s_j) : s_j \in SP_j^k(t_j)\}$, we have:

$$SP_i^{k+1}(t_i) := \left\{ s_i \in RP_i^0(t_i) : \exists \mu_i \in \Delta(SP_j^k) \cap C_i(t_i) \text{ s.t. } BR_i(\mu_i; t_i) = \{s_i\} \right\}$$

Finally, set $SP_i(t_i) := \bigcap_{k \geq 0} SP_i(t_i)$. These auxiliary solution concept will be useful in the proof of both statements of the theorem. The counterpart of $SP_i$ for type spaces $\mathcal{T}$ that are not the universal one, $SP_i^{\mathcal{T}}$, is defined introducing the obvious, necessary changes.

## C.1 Theorem 1, Part (*i*)

**Proof of part (*i*) of Theorem 1.** We proceed by induction on $k$. The claim holds trivially for the case $k = 0$. Let's suppose that so does it for $k \geq 0$; we check next that then, the claim also holds for $k+1$. Fix player $i$ and types $t_i \in T_i^*$ and $t_i' \in \Lambda_i^{k+1}(t_i)$. Clearly, it suffices to show only one inclusion. Pick $s_i \in SP_i^{k+1}(t_i)$ and $\mu_i \in \Delta(SP_j^k) \cap C_i(t_i)$ such that $BR_i(\mu_i; t_i) = \{s_i\}$. Now, define measure $\nu_i \in \Delta(T_j \times S_j)$ as follows: $\nu_i[E] = \tau_i[\mathrm{Proj}_{T_j}(E)]$ for any measurable $E \subseteq T_j^* \times S_j$. Clearly, $\mu_i$ is absolutely continuous w.r.t. to $\nu_i$, and therefore, we can pick the corresponding Radon-Nykodym derivative, which we denote by $f_i$. Then, define measure $\mu_i' \in \Delta(T_j^* \times S_j)$ by setting $\mu_i'[E] := \int_E f_i \mathrm{d}\nu_i'$ for any measurable $E \subseteq T_j^* \times S_j$, where $\nu_i'[E] := \tau_i(t_i')[\mathrm{Proj}_{T_j^*}(E)]$. Clearly, $\mu_i' \in C_i(t_i')$ and $BR_i(\mu_i'; t_i')$. In addition, the fact that $t_i' \in \Lambda_i^{k+1}(t_i)$, together with the induction hypothesis guarantees that $\mu_i' \in \Delta(SP_j^k)$. Hence, we conclude that $s_i \in SP_i^{k+1}(t_i')$. ■

## C.2 Theorem 1, Part *(ii)*

We present first two auxiliary results and proceed next with the proof of the claim.

**Lemma 8** *Correspondence $RP_i^k : T_i^* \rightrightarrows S_i$ is upper-hemicontinuous for any $k \geq 0$.*

**Proof.** The claim is true for $\theta_i''$-types: the correspondences are by definition constant around them. Hence, we only need to complete the proof for $\theta_i'$-types. We proceed by induction on $k$. The initial step ($k = 0$) is trivially true, so let's move on to the inductive one. Suppose that $k \geq 0$ is such that the claim holds; we are going to check that it is also true for $k+1$. To see it, fix player $i$ and take convergent sequence of $\theta_i'$-types $(t_i^\nu)_{\nu \in \mathbb{N}}$ with limit $t_i$ (which is thus a $\theta_i'$-type too) and action $a_i \in A_i$ such that $a_i \in RP_i^{k+1}(t_i^\nu)$ for any $\nu \in \mathbb{N}$. For each $\nu \in \mathbb{N}$ take conjecture $\mu_i^\nu \in \Delta(RP_j^k) \cap C_i(t_i^\nu)$ satisfying that $a_i \in BR_i(\mu_i^\nu; t_i^\nu)$. We know from compactness of $\Delta(RP_j^0)$ that there exists some convergent subsequence $(\mu_i^{\nu_m})_{m \in \mathbb{N}} \subseteq (\mu_i^\nu)_{\nu \in \mathbb{N}}$ whose limit we denote by $\mu_i$. Continuity of taking marginals guarantees that $\mu_i \in C_i(t_i)$, and we know due to the upper-hemicontinuity of $BR_i$ that $a_i \in BR_i(\mu_i; t_i)$. Now, according to the induction hypothesis, $RP_j^k(\cdot)$ is upper-hemicontinuous, and then, it follows from the Closed Graph Theorem that $RP_j^k$ is closed. It follows that:

$$\mu_i[R_j^k] \geq \underset{m \to \infty}{\text{limsup}} \mu_i^{\nu_m}[RP_j^k] = 1.$$

That is, $\mu_i \in \Delta(RP_j^k)$. This way, we conclude that $a_i \in RP_i^{k+1}(t_i)$, and thus, that $RP_i^{k+1}(\cdot)$ is upper-hemicontinuous. ■

**Lemma 9** *For any finite type $t_i$ and for any $k \geq 0$ there exists a convergent sequence of finite types $(t_i^{k,\nu})_{\nu \in \mathbb{N}}$ with limit $t_i$ and such that $SP_i^k(t_i^{k,\nu}) \neq \emptyset$ for any $\nu \in \mathbb{N}$.*

**Proof.** The claim holds trivially for $\theta_i''$-types, so let's complete the proof for the $\theta_i'$-types. We proceed by induction on $k$. The initial step ($k = 0$) is immediate: $SP_i^0(t_i) = \mathcal{B}_i$ for any player $i$ and any $\theta_i'$-type $t_i$. For the inductive step, suppose that $k \geq 0$ is such that for any player $i$ and any finite $\theta_i'$-type $t_i$ there exists a convergent sequence of finite types, $\{t_i^{k,\nu}\}_{\nu \geq 0}$, with limit $t_i$ such that $SP_i^k(t_i^\nu) \neq \emptyset$ for any $k \in \mathbb{N}$. We check next that the claim also holds for $k+1$. To see it, fix player $i$ and finite $\theta_i'$-type $t_i$. If $SP_i^{k+1}(t_i) \neq \emptyset$, then the claim holds trivially: simply set $t_i^\nu = t_i$ for any $\nu \in \mathbb{N}$; so let's suppose that $SP_i^{k+1}(t_i) = \emptyset$. We know from the induction hypothesis that for any $t_j$ in the (finite) support of $\tau_i^*(t_i)$ there exists a sequence of finite types $(t_j^\nu)_{\nu \in \mathbb{N}}$ with limit $t_j$ such that $SP_j^k(t_j^\nu) \neq \emptyset$. Define then, for any $\nu \in \mathbb{N}$, belief $\tau_i^\nu \in \Delta(T_j^*)$ as follows:

$$\tau_i^\nu[E] := \tau_i(t_i)\left[\left\{t_j \in T_j^* : t_j^\nu \in E\right\}\right],$$

for any measurable $E \subseteq T_j^*$. The fact that $\tau_i^*(t_i)$ has finite support guarantees that every $\tau_i^\nu$ is well-defined. Set now, for each $\nu \in \mathbb{N}$, type $\bar{t}_i^\nu = \tau_i^{*,-1}(\tau_i^\nu)$. Clearly, every type $\bar{t}_i^\nu$ is finite and $(\bar{t}_i^\nu)_{\nu \in \mathbb{N}}$ converges to $t_i$.[29] In addition, we know that, for every $\nu \in \mathbb{N}$,

[29]It is easier to notice that $\{\tau_i^\nu\}_{\nu \in \mathbb{N}}$ converges to $\tau_i^*(t_i)$: pick open set $U \subseteq T_j^*$ and notice that since $\tau_i^*(t_i)$ is

by construction, $SP_j^k(t_j) \neq \emptyset$ for every $t_j \in \text{supp}\tau_i(\bar{t}_i^\nu)$, and thus, that set of admissible conjectures $C_i^k(\bar{t}_i^\nu) = \{\mu_i \in \Delta(SP_j^k) : \text{marg}_{T_j}\mu_i = \tau_i(t_i^\nu)\}$ is non-empty. Then, if for any $\nu \in \mathbb{N}$ there exists some $\mu_i^\nu \in C_i^k(\bar{t}_i^\nu)$ such that $|BR_i(\mu_i^\nu; \bar{t}_i^\nu)| = 1$, the proof is complete: $(\bar{t}_i^\nu)_{\nu\in\mathbb{N}}$ is a sequence converging to $t_i$ and satisfying that $SP_i^{k+1}(\bar{t}_i^\nu) \neq \emptyset$ for every $\nu \in \mathbb{N}$. If there exists some $\nu \in \mathbb{N}$ such that for any $\mu_i^\nu \in C_i^k(\bar{t}_i^\nu)$, $|BR_i(\mu_i^\nu; \bar{t}_i^\nu)| \geq 2$, fix such that $\nu$ and arbitrary $\mu_i^\nu \in C_i^k(\bar{t}_i^\nu)$, and denote:

$$\{a_i^0, \ldots, a_i^N\} = BR_i(\mu_i^\nu; t_i^\nu).$$

By Assumption 1, we can assume w.l.o.g. that,

$$\{a_i^0\} = \arg\max_{\ell=0,\ldots,N} u_i^*(a_i^\ell, a_j^*(a_i^\ell)).$$

Then, notice that if $\bar{t}_i^\nu = (\theta_i', \pi_i)$ and we denote $p = \pi_{i,1}[\omega^0]$ and,

$$\hat{\mu}_i[a_j] := \left(\frac{1}{p}\right) \cdot \mu_i[T_j(\theta_j') \times \{a_j\}],$$

for any $a_j \in A_j$, then there exists some $\varepsilon > 0$ such that for any $m \in \mathbb{N}$,

$$\{a_j^0\} = \arg\max_{a_i \in A_i} \left( \left(p - \frac{\varepsilon}{m}\right) \cdot \sum_{a_j \in A_j} \hat{\mu}_i[a_j] \cdot u_i^*(a_i, a_j) + \left(1 - p + \frac{\varepsilon}{m}\right) \cdot u_i^*(a_i, a_j^*(a_i)) \right).$$

For each $m \in \mathbb{N}$ define now belief $\mu_i^{\nu,m} \in \Delta(R_j^0)$ as follows:

$$\mu_i^{\nu,m}[(t_j, s_j)] := \begin{cases} \left(\frac{p-\varepsilon/m}{p}\right) \cdot \mu_i[(t_j, s_j)] & \text{if} t_j \in T_j(\theta_j'), \\ \left(\frac{1-p+\varepsilon/m}{1-p}\right) \cdot \mu_i[(t_j, s_j)] & \text{if} t_j \in T_j(\theta_j''). \end{cases}$$

and set (finite) type $t_i^{\nu,m} = \tau_i^{*,-1}(\mu_i^{\nu,m})$. Clearly, the following three hold for any $m \in \mathbb{N}$: $(i)$ $\{a_i^0\} = BR_i(\mu_i^{\nu,m}; t_i^{\nu,m})$, $(ii)$ $\mu_i^{\nu,m} \in \Delta(SP_j^k)$ (because $\mu_i^{\nu,m}$ and $\mu_i^\nu$ only diverge at first order beliefs) and $(iii)$ $a_i^0 \in \mathcal{B}_i$ (because $\hat{\mu}_i \in \Delta(\mathcal{B}_j)$). Thus, we conclude that $a_i^0 \in SP_i^{k+1}(t_i^{\nu,m})$ for every $m \in \mathbb{N}$. Denote now $t_i^\nu = t_i^{\nu,\nu}$ for every $\nu \in \mathbb{N}$. Then $(t_i^\nu)_{\nu\in\mathbb{N}}$ is a sequence of finite types that converges to $t_i$ and such that $SP_i^{k+1}(t_i^\nu) \neq \emptyset$ for every $\nu \in \mathbb{N}$. ■

**Proof of part *(ii)* of Theorem 1.** Upper-hemicontinuity follows from the fact that each $RP_i^k$ is upper-hemicontinuous. For non-emptiness pick some arbitrary finite type $t_i$. We know from the previous lemma that for any $k \geq 0$ there exists a sequence of finite types $(t_i^{k,\nu})_{\nu\in\mathbb{N}}$ with limit $t_i$ and such that $SP_i^k(t_i^{k,\nu}) \neq \emptyset$ for any $\nu \in \mathbb{N}$. Since $SP_i^k(\cdot)$ is a refinement of $RP_i^k(\cdot)$ we know that $RP_i^k(t_i^{k,\nu}) \neq \emptyset$ for any $k \geq 0$ and any $\nu \in \mathbb{N}$. Thus, it follows from the upper-hemicontinuity of each $RP_i^k(\cdot)$ that $RP_i^k(t_i) \neq \emptyset$, and hence, that $RP_i(t_i) \neq \emptyset$. Then, the upper-hemicontinuity of $RP_i$ and the fact that the set of finite types is dense implies that

finite, there exists some $\nu_0 \in \mathbb{N}$ such that for any $\nu \geq \nu_0$, $\{t_j^{\nu,1}, \ldots, \tau_j^{\nu,N}\} \subseteq U$ for $\{t_j^1, \ldots, t_j^N\} = U \cap \text{supp}\tau_i^*(t_i)$. Then, clearly, $\tau_i^\nu[U] \geq \tau_i^*(t_i)[U]$ for any $\nu \geq \nu_0$, and hence, $\tau_i^*(t_i)[U] \leq \text{liminf}_{\nu\to\infty}\tau_i^\nu[U]$.

$RP_i$ is non-empty. ■

### C.3 Theorem 1, Part *(iii)*

We begin proving in Lemma 10 that for every $k \in \mathbb{N}$ and every action $a_i \in \mathcal{B}_i^k$ there exists some finite $\theta_i'$-type $t_i^{a_i}$ such that $SP_i^{k+1}(t_i^{a_i}) = SP_i(t_i^{a_i}) = \{a_i\}$. Based in this result we create two different kinds of successive perturbations. First, Lemma 11 shows that for any finite type $t_i$ and any strategy $s_i \in RP_i\,(t_i)$, there exists some finite type $t_i'$ arbitrarily close to $t_i$ and such that $s_i \in SP_i(t_i')$. Second, Lemma 12 shows that for any finite type $t_i$ and any strategy $s_i \in SP_i(t_i)$, there exists some finite type $t_i'$ arbitrarily close to $t_i$ and such that $R_i(t_i') = SP_i(t_i') = \{s_i\}$. Then, we finally present the proof of part $(iii)$ of Theorem 1.

**Lemma 10** *For any $k \in \mathbb{N}$ and $a_i \in \mathcal{B}_i^k$, there exists a finite type $t_i^{a_i} \in T_i^*$ such that $R_i^{k+1}(t_i^{a_i}) = SP_i(t_i^{a_i}) = \{a_i\}$.*

**Proof.** We proceed by induction on $k$. For the initial step ($k = 1$) it is clear that under Assumption 1 $R_i^2(t_i^{CB}(\omega^i)) = SP_i(t_i^{a_i}) = \{a_i^i\}$. For the inductive step suppose that the statement is true for all $n = 1, ..., k$: i.e., for each $n = 1, ..., k$, and for each $a_i \in \mathcal{B}_i^n$ there exists some finite type $t_i^{a_i} \in T_i^*$ such that $R_i^{n+1}(t_i^{a_i}) = SP_i(t_i^{a_i}) = \{a_i\}$. Then, let $a_i \in \mathcal{B}_i^{k+1}$. By definition, there exists some conjecture $\mu_i \in \Delta(\mathcal{B}_j^k)$ and some $p \in [0,1]$ such that:

$$\{a_i\} = \arg\max_{a_i' \in A_i} \left( p \cdot \sum_{a_j \in A_j} \mu_i[a_j] \cdot u_i^*(a_i', a_j) + (1-p) \cdot u_i^*(a_i', a_j^*(a_i')) \right).$$

Under the inductive hypothesis, for each $a_j \in \operatorname{supp} \mu_i$ there exists some finite type $t_j^{a_j}$ such that $R_i^k(t_j^{a_j}) = SP_i(t_j^{a_j}) = \{a_j\}$. Hence, let $t_j'$ be such that $\omega^i \in \hat{\theta}_j\,(t_i')$, and set $t_i^{a_i} := \tau_i^{*,-1}(\tau_i)$ where:

$$\tau_i\left[(\omega, t_j)\right] := \begin{cases} p \cdot \mu[a_j] & \text{if } (t_j, \omega) = (\omega^0, t_j^{a_j}), \\ 1-p & \text{if } (t_j, \omega) = (\omega^i, t_j'). \end{cases}$$

By construction, $R_i^{k+1}\,(t_i^{a_i}) = SP_i(t_i^{a_i}) = \{a_i\}$. ■

The 'dominance types' obtained in the previous lemma allow for the first desired class of perturbation, which approximates the original type $t_i$ by some arbitrarily close type $t_i'$ so that $s_i \in SP_i(t_i')$ for any previously fixed $s_i \in RP_i\,(t_i)$.

**Lemma 11** *For any finite type $t_i$ and any $s_i \in RP_i(t_i)$ there exists a sequence of finite types $(t_i^\nu)_{\nu \in \mathbb{N}}$ converging to $t_i$ and such that $s_i \in SP_i(t_i^\nu)$ for any $\nu \in \mathbb{N}$.*

**Proof.** During the proof, for a given type structure $\mathcal{T}$ and player $i$'s type $t_i$ we write $RP_i(t_i)$ to refer to set $RP_i(\phi_i(t_i))$.[30] Then, the claim is satisfied trivially if $t_i$ is a $\theta_i''$-type, so let's complete the proof for the $\theta_i'$-types. Fix player $i$ and finite $\theta_i'$-type $\hat{t}_i \in T_i^*$. Since $\hat{t}_i$ is finite

[30] We skip the details for the sake of expositional brevity. However, the abuse of notation is justified by the fact that it is immediate to define a correspondence $RP_i^{\mathcal{T}} : T_i^* \rightrightarrows S_i$ capturing the same intuition behind $RP_i$, and check, analogously as done in Lemma 1, that type space invariance is satisfied.

we know that there exists some finite type structure $\mathcal{T} = (T_i, \hat{\theta}_i, \tau_i)_{i\in I}$ such that $\phi_i(\bar{t}_i) = \hat{t}_i$ for some $\bar{t}_i \in T_i$. Then, for each player $i$ and each $n \in \mathbb{N}$ define now the following set of types:

$$T_i^n := \{n\} \times \{(t_i, a_i) \in T_i \times A_i : a_i \in RP_i(t_i)\} \cup \{(t_i^{a_i}, a_i) : a_i \in \mathcal{B}_i\},$$

where for each $a_i \in \mathcal{B}_i$ type $t_i^{a_i}$ is, as obtained in Lemma 10, a finite type satisfying that $SP_i(t_i^{a_i}) = \{a_i\}$. Now, for each player $i$ and each $(t_i, a_i) \in T_i \times RP_i(t_i)$ fix some:[31]

$$\mu_i(t_i, a_i) \in \Delta(RP_j) \cap C_i(t_i) \text{ s.t. } a_i \in BR_i(\mu_i(t_i, a_i); t_i),$$

and for each $a_i \in \mathcal{B}$,[32]

$$\mu_i^{a_i} \in \Delta(SP_j) \cap C_i(t_i^{a_i}),$$

which must satisfy by construction of $t_i^{a_i}$ that $BR_i(\mu_i^{a_i}; t_i^{a_i}) = \{a_i\}$. Then, for each player $i$, each $n \in \mathbb{N}$ and each $(n, t_i, a_i) \in T_i^n$ set first,

$$\gamma_i^n(t_i, a_i) := \begin{cases} (n, t_i, a_i) & \text{if } (t_i, a_i) \in T_i \times RP_i(t_i), \\ (n, t_i^{a_i}, a_i) & \text{otherwise,} \end{cases}$$

and second,

$$\tau_i^n(n, t_i, a_i) := \begin{cases} \left(1 - \frac{1}{n}\right) \cdot \mu_i(t_i, a_i) \circ \gamma_j^{n,-1} + \left(\frac{1}{n}\right) \cdot \mu_i^{a_i} \circ \gamma_j^{n,-1} & \text{if } (t_i, a_i) \in T_i \times RP_i(t_i), \\ \mu_i^{a_i} \circ \gamma_j^{n,-1} & \text{otherwise.} \end{cases}$$

Finiteness of $T_i^n$ and $T_j^n$ guarantees that $\tau_i^n : T_i^n \to \Delta(T_j^n)$ is well-defined and continuous. Finally, set $\hat{\theta}_i^n(n, t_i, a_i) = \hat{\theta}_i(t_i)$. Obviously, $\mathcal{T}^n := (T_i^n, \hat{\theta}_i^n, \tau_i^n)_{i\in I}$ is a well-defined type space for any $n \in \mathbb{N}$. Then, we have that:

- For any player $i$ and any pair $(t_i, a_i) \in T_i \times RP_i(t_i)$, sequence of belief hierarchies $(\hat{\pi}_i(\nu, t_i, a_i))_{\nu\in\mathbb{N}}$ converges to $\hat{\pi}_i(t_i)$.
- For any player $i$, any pair $(t_i, a_i) \in T_i \times RP_i(t_i)$ and any $\nu \in \mathbb{N}$ pick $\mu_i(t_i, a_i)$ and define $\mu_i^\nu(t_i, a_i)$ as follows:

$$\mu_i^\nu(t_i, a_i) := \left(1 - \frac{1}{\nu}\right) \cdot \mu_i(t_i, a_i) \circ \Gamma_j^{\nu,-1} + \left(\frac{1}{\nu}\right) \cdot \mu_i^{a_i} \circ \Gamma_j^{\nu,-1},$$

where,

$$\Gamma_j^\nu(t_j, a_j) := \begin{cases} (\phi_j(\nu, t_j, a_j), a_j) & \text{if } (t_j, a_j) \in T_j \times RP_j(t_j), \\ (\phi_j((\nu, t_j^{a_j}, a_j), a_j) & \text{otherwise.} \end{cases}$$

Clearly, the following two properties hold by construction: $\mu_i^\nu(t_i, a_i) \in C_i(\phi_i(\nu, t_i, a_i))$

[31] For the existence of the following conjecture, notice that since $\mathcal{T}$ is finite we know that there exists some $N \in \mathbb{N}$ such that $RP_i(t_i) = RP_i^N(t_i)$ for any $t_i \in \mathcal{T}$ and any player $i$. In particular, this implies that $\bigcap_{k\geq 0} \Delta(RP_i^k) = \bigcap_{k=0}^N \Delta(RP_i^k)$, and thus, that we can think of each $RP_i^k(t_i)$ as the set of best replies to conjectures that induce $t_i$ and put probability 1 on $\Delta(RP_j)$.

[32] For the existence of the following conjecture, notice that, similarly as discussed in the previous section, the fact that for each player $i$ set $\mathcal{B}$ is finite and so are types $t_i^{a_i}$ where $a_i \in \mathcal{B}_i$, guarantees that there exists some $M \in \mathbb{N}$ such that $\mu_i \in C_i(t_i^{a_i})$ and $\Delta(SP_j)$ if and only if $\mu_i \in C_i(t_i^{a_i})$ and in $\bigcap_{k=0}^M \Delta(SP_j^k)$.

and $BR_i(\mu_i^\nu(t_i,a_i);\phi_i(\nu,t_i,a_i))=\{a_i\}$.

- It follows then that for any player $i$ and any pair $(t_i,a_i)\in T_i\times RP_i(t_i)\cup\{(t_i^{a_i},a_i)\}$,

$$a_i\in SP_i^1(\phi_i(\nu,t_i,a_i)).$$

Suppose now that $k\geq 1$ is such that we have that for any player $i$ and any pair $(t_i,a_i)\in(T_i\times RP_i(t_i))\cup\{(t_i^{a_i},a_i)\}$,

$$a_i\in SP_i^k(\phi_i(\nu,t_i,a_i)).$$

Fix then player $i$, pair $(t_i,a_i)\in(T_i\times RP_i(t_i))\cup\{(t_i^{a_i},a_i)\}$ and pick $(\mu_i^\nu(t_i,a_i))_{\nu\in\mathbb{N}}$ as constructed above. We already know from the previous bullet that $\mu_i^\nu(t_i,a_i)\in C_i(\phi_i(\nu,t_i,a_i))$ and $BR_i(\mu_i^\nu(t_i,a_i);\phi_i(\nu,t_i,a_i))=\{a_i\}$. Note in addition that each finite $\phi_i(\nu,t_i,a_i)$ only assigns positive probability to finite belief hierarchies $\phi_j(\nu,t_j,a_j)$ where $(t_j,a_j)\in(T_j\times RP_j(t_j))\cup\{(t_j^{a_j},a_j)\}$. Then, we know that $\mu_i^\nu(t_i,a_i)$ only assigns positive probability to pairs,

$$(\phi_j(\nu,t_j,a_j),a_j)\text{ such that }(t_j,a_j)\in(T_j\times RP_j(t_j))\cup\{(t_j^{a_j},a_j)\}$$

Hence, we conclude that $\mu_i^\nu(t_i,a_i)$ assigns full probability to $SP_j^k$, and therefore, that,

$$a_i\in SP_i^{k+1}(\phi_i(\nu,t_i,a_i)).$$

Thus, it follows by induction that for any player $i$, $a_i\in SP_i(\phi_i(\nu,t_i,a_i))$ for any pair $(t_i,a_i)\in(T_i\times RP_i(t_i))\cup\{(t_i^{a_i},a_i)\}$ and any $\nu\in\mathbb{N}$.

Then, given the above, fix $\hat{a}_i\in RP_i(\hat{t}_i)$ and pick sequence $(t_i^\nu)_{\nu\in\mathbb{N}}$ setting $t_i^\nu:=\phi_i(\nu,\hat{t}_i,\hat{a}_i)$ for any $\nu\in\mathbb{N}$. As seen above, $(t_i^\nu)_{\nu\in\mathbb{N}}$ is a sequence of finite types converging to $\hat{t}_i$ and such that $\hat{a}_i\in SP_i(t_i^\nu)$ for any $\nu\in\mathbb{N}$. ■

The following lemma relies again in Lemma 10 to perturb a type $t_i$ so that any strategy in $s_i\in SP_i(t_i)$ can be made uniquely rationalizable for some $t_i'$ arbitrarily close to $t_i$.

**Lemma 12** *For any player $i$, any $k\geq 1$, any finite type $t_i=(\theta_i,\pi_i)$ and any strategy $s_i\in SP_i^{m+k}(t_i)$ there exists some finite type $t_i^k=(\theta_i,\pi_i^k)$ such that (i) $\pi_{i,k}^k=\pi_{i,k}$ and (ii) $R_i^{m+k+1}(t_i^k)=SP_i(t_i^k)=\{s_i\}$.*

**Proof.** Throughout the proof, for any player $i$ we denote $T_i'=\{\theta_i'\}\times H_i(\theta_i')$ and $T_i''=\{\theta_i''\}\times H_i(\theta_i'')$. Notice now that the result is immediately true for $\theta_i''$-types; thus we can focus on the $\theta_i'$-types. Let's proceed by induction on $k$:

*Initial Step* ($k=1$): Fix player $i$, finite $\theta_i'$-type $t_i=(\theta_i,\pi_i)$, action $a_i\in SP_i^{m+1}(t_i)$ and conjecture $\mu_i\in\Delta(SP_j^m)$ such that $BR_i(\mu_i;t_i)=\{a_i\}$. We know from Lemma 10 that for any action $a_j\in S_j(\omega^0)$ such that $\mu_i[T_j'\times\{a_j\}]>0$ there exists some finite type $t_j(a_j)$ such that $SP_j^{m+1}(t_j(a_j)) = \{a_j\}$. Define then conjecture $\hat{\mu}_i \in \Delta(T_j \times S_j)$ as follows:

$$\hat{\mu}_i[(t_j, s_j)] := \begin{cases} \mu_i[T_j' \times \{s_j\}] & \text{if } s_j \in S_j(\omega^0) \text{ and } t_j = t_j(s_j), \\ \mu_i[(t_j, s_j)] & \text{if } t_j \in T_j''), \\ 0 & \text{otherwise,} \end{cases}$$

for any $(t_j, s_j) \in T_j \times S_j$. Finiteness of $t_i$ guarantees that the above extends to a unique, well-defined measure. Set now $t_i^1 := \tau_i^{-1}(\mathrm{marg}_{T_j}\hat{\mu}_i)$. Obviously, $t_i^1 = (\theta_i', \pi_i^1)$ is finite and satisfies that $\pi_{i,1}^1 = \pi_i^1$. Now, pick arbitrary conjecture $\check{\mu}_i \in \Delta(R_j^{m+1})$ such that $\mathrm{marg}_{T_j^*}\check{\mu}_i = \tau_i^*(t_i^1)$. We need to check that $BR_i(\check{\mu}_i; t_i^1) = \{a_i\}$. To see it, notice that the following two hold for any $a_j \in S_j(\omega^0)$ and any $s_j \in S_j(\omega^i)$:

$$\check{\mu}_i[T_j' \times \{a_j\}] = \check{\mu}_i[(t_j(a_j), a_j)] \text{ and } \check{\mu}_i[T_j' \times \{s_j\}] = \mu_i[T_j' \times \{s_j\}].$$

Obviously, this implies that,

$$\int_{(t_j,s_j)} u_i(s_i, s_j, \omega(t_i^1, t_j))\mathrm{d}\check{\mu}_i = \int_{(t_j,s_j)} u_i(s_i, s_j, \omega(t_i, t_j))\mathrm{d}\mu_i,$$

and hence, we conclude that $BR_i(\check{\mu}_i; t_i^1) = BR_i(\mu_i; t_i)$. In consequence, we have $R_i^{m+2}(t_i^1) = \{a_i\}$. Now, remember that we also know from Lemma 10 that $SP_j(t_j(a_j)) = \{a_j\}$ for any $a_j \in S_j(\omega^0)$ such that $\mu_i[T_j' \times \{a_j\}] > 0$. Since within $T_j' \times S_j(\omega^0)$ conjecture $\hat{\mu}_i$ puts full probability on set $\{(t_j(a_j), a_j) : \mu_i[T_j' \times \{a_j\}] > 0\}$, it follows then that $\hat{\mu}_i \in \Delta(SP_j)$ (and notice that $\Delta(SP_j) \subseteq \bigcap_{n\geq 0}\Delta(SP_j^n)$). Finally, the fact that, as seen above, $BR_i(\hat{\mu}_i; t_i^1) = \{a_i\}$ lets us additionally conclude that $SP_i(t_i^1) = \{a_i\}$.

*Inductive Step*: Suppose now that $k \geq 1$ is such that the claim holds. We check next that so does for $k+1$. To see it, fix player $i$, finite $\theta_i'$-type $t_i$, action $a_i \in SP_i^{m+(k+1)}(t_i)$ and conjecture $\mu_i \in \Delta(SP_j^{m+k})$ such that $BR_i(\mu_i; t_i) = \{a_i\}$. Then, we know from the induction hypothesis that for every $(t_j, s_j) \in \mathrm{supp}\ \mu_i$ such that $t_j = (\theta_j', \pi_j)$ there exists some finite $t_j^k(t_j, s_j) = (\theta_j', \pi_j^k(t_j, s_j)) \in T_j$ such that $\pi_{j,k}^k(t_j, s_j) = \pi_{j,k}$ and,

$$R_j^{m+(k+1)}(t_j^k(t_j, s_j)) = SP_j(t_j^k(t_j, s_j)) = \{s_j\}.$$

Define then conjecture $\hat{\mu}_i \in \Delta(R_j^0)$ as follows:

$$\hat{\mu}_i[(t_j, s_j)] := \mu_i[\{\hat{t}_j \in T_j : t_j = t_j^k(\hat{t}_j, s_j)\} \times \{s_j\}],$$

for any $(t_j, s_j) \in R_j^0$. The fact that $t_i$ is finite guarantees that $\hat{\mu}_i$ is a well-defined element of $\Delta(R_j^0)$. Then, set $t_i^{k+1} := \tau_i^{-1}(\mathrm{marg}_{T_j}\hat{\mu}_i)$ and denote $t_i^{k+1} = (\theta_i', \pi_i^{k+1})$. Note that the following two properties hold:

(*i*) $\pi_{i,k+1}^{k+1} = \pi_{i,k+1}$. Pick arbitrary $(\omega, \pi_{j,1}, \dots, \pi_{j,k}) \in \Omega \times Z_j^k$. Then:

$$\pi_{i,k+1}^{k+1}[(\omega, \pi_{j,1}, \dots, \pi_{j,k})] =$$

$$= \tau_i^*(t_i^{k+1}) \left[ t_j \in T_j : \begin{array}{ll} (i) & \omega \in \theta_j(t_j) \cap \theta_i(t_i), \\ (ii) & \hat{\pi}_{j,\ell}(t_j) = \pi_{j,\ell} \text{ for any } \ell = 1, \ldots, k \end{array} \right]$$

$$= \hat{\mu}_i \left[ \left\{ t_j \in T_j : \begin{array}{ll} (i) & \omega \in \theta_j(t_j) \cap \theta_i(t_i), \\ (ii) & \hat{\pi}_{j,\ell}(t_j) = \pi_{j,\ell} \text{ for any } \ell = 1, \ldots, k \end{array} \right\} \times S_j \right]$$

$$= \mu_i \left[ \left\{ t_j \in T_j : \begin{array}{ll} \multicolumn{2}{l}{\text{There is some } (\check{t}_j, s_j) \in T_j \times S_j \text{such that:}} \\ (i) & \check{t}_j = t_j^k(t_j, s_j), \\ (ii) & \omega \in \theta_j(\check{t}_j) \cap \theta_i(t_i), \\ (iii) & \hat{\pi}_{j,\ell}(\check{t}_j) = \pi_{j,\ell} \text{ for any } \ell = 1, \ldots, k \end{array} \right\} \times S_j \right]$$

$$= \mu_i \left[ \left\{ t_j \in T_j : \begin{array}{ll} (i) & \omega \in \theta_j(t_j) \cap \theta_i(t_i), \\ (ii) & \hat{\pi}_{j,\ell}(t_j) = \pi_{j,\ell} \text{ for any } \ell = 1, \ldots, k \end{array} \right\} \times S_j \right]$$

$$= \pi_{i,k+1}[(\omega, \pi_{j,1}, \ldots, \pi_{j,k})].$$

Thus, the claim follows from finiteness of $t_i$.

(*ii*) $BR_i(\check{\mu}_i; t_i^{k+1}) = \{a_i\}$ for any $\check{\mu}_i \in \Delta(R_j^{m+(k+1)})$ such that $\mathrm{marg}_{T_j}\check{\mu}_i = \tau_i(t_i^{k+1})$. To see it, notice that the following two hold for any $a_j \in S_j(\omega^0)$ and any $s_j \in S_j(\omega^i)$:

$$\check{\mu}_i[T_j' \times \{a_j\}] = \check{\mu}_i[\{t_j^k(t_j, a_j) : (t_j, a_j) \in T_j' \times S_j \cap R_j^{m+(k+1)}\} \times \{a_j\}] = \mu_i[T_j' \times \{a_j\}],$$
$$\check{\mu}_i[T_j'' \times \{s_j\}] = \mu_i[T_j'' \times \{s_j\}].$$

Obviously, this implies that,

$$\int_{(t_j,s_j)} u_i(s_i, s_j, \omega(t_i^{k+1}, t_j))\mathrm{d}\check{\mu}_i = \int_{(t_j,s_j)} u_i(s_i, s_j, \omega(t_i, t_j))\mathrm{d}\mu_i,$$

and hence, we conclude that $BR_i(\check{\mu}_i; t_i^{k+1}) = BR_i(\mu_i; t_i) = \{a_i\}$.

Thus, it follows from (*i*) that $R_i^{m+k+2}(t_i^{k+1}) = \{a_i\}$. Now, remember that we also knew from the induction hypothesis that $SP_j(t_j^k(t_j, a_j)) = \{a_j\}$ for any $(t_j, a_j) \in T_j' \times S_j(\omega^0)$ such that $\mu_i[(t_j, a_j)] > 0$. Since within $T_j' \times S_j(\omega^0)$ conjecture $\hat{\mu}_i$ puts full probability on set $\{(t_j^k(t_j, a_j), a_j) : \mu_i[(t_j, a_j)] > 0\}$, it follows then that $\hat{\mu}_i \in \Delta(SP_j)$. Then, the fact that, as seen above, $BR_i(\hat{\mu}_i; t_i^{k+1}) = \{a_i\}$ lets us additionally conclude that $SP_i(t_i^{k+1}) = \{a_i\}$. ■

**Proof of part (*iii*) of Theorem 1.** Fix player $i$, finite type $t_i$ and strategy $s_i \in RP_i(t_i)$. Then, according to Lemma 11 there exists some convergent sequence of finite types $(\bar{t}_i^\nu)_{\nu\in\mathbb{N}}$ with limit $t_i$ such that $s_i \in SP_i^\nu(\bar{t}_i^\nu)$ for any $\nu \in \mathbb{N}$. Then, we know from Lemma 12 that for any $\nu \in \mathbb{N}$ there exists a sequence of finite types $(t_i^{\nu,k})_{k\in\mathbb{N}}$ such that $R_i^{m+\nu+1}(t_i^{\nu,\nu}) = SP_i(t_i^{\nu,\nu}) = \{s_i\}$ and $\pi_{i,\nu}(t_i^{\nu,\nu}) = \pi_{i,\nu}(\bar{t}_i^\nu)$ for any $\nu \in \mathbb{N}$. Notice finally that $SP_i$ refines $RP_i$ and the latter refines $R_i$. Then, sequence $(t_i^\nu)_{\nu\in\mathbb{N}}$ where $t_i^\nu = t_i^{\nu,\nu}$ for any $\nu \in \mathbb{N}$ consists of finite types, converges to $t_i$, and satisfies that $R_i(t_i^\nu) = RP_i(t_i^\nu) = \{s_i\}$ for any $\nu \in \mathbb{N}$. ■

### C.4 Corollaries 1, 2 and 3

**Proof of Corollary 1.** Suppose that $t_i$ is a finite type for which there exist some $s_i \in R_i(t_i)$ and some sequence of finite types $(t_i^\nu)_{\nu\in\mathbb{N}}$ such that $R_i(t_i^\nu) = \{s_i\}$. Then, since $RP_i$ is a non-empty refinement of $R_i$ we know that $RP_i\,(t_i^\nu) = \{s_i\}$ for any $\nu \in \mathbb{N}$. Hence, it follows from upper-hemicontinuity of $RP_i$ that $s_i \in RP_i\,(t_i)$. ■

**Proof of Corollary 2.** Upper hemicontinuity of both $R_i$ and $RP_i$ implies that the following set is open: $U_i = \{t_i \in T_i^* : R_i(t_i) = RP_i\,(t_i) \text{ and } |R_i(t_i)| = 1\}$. It follows from part *(iii)* of Theorem 1 that it is, in addition, dense. ■

**Proof of Corollary 3.** Suppose that $W_i : T_i^* \rightrightarrows S_i$ is an upper-hemicontinuous and never empty-valued refinement of $RP_i$. Then, there exists some finite $t_i \in T_i^*$ and some $s_i \in RP_i(t_i) \setminus W_i(t_i)$. Now, we know by Theorem 1 that there exists some sequence $(t_i^\nu)_{\nu\in\mathbb{N}}$ converging to $t_i$ such that $RP_i(t_i) = \{s_i\}$ for any $\nu \in \mathbb{N}$. Since $W_i$ is a refinement, $W_i(t_i^\nu) = \{s_i\}$ for any $\nu \in \mathbb{N}$, and hence, upper-hemicontinuity implies that $s_i \in W_i(t_i)$. Thus, we reached a contradiction. ■

## D Robust Refinements

**Proof of Proposition 1.** W.l.o.g. in an agreement game with uniform disagreement payoffs, let $u_i^*(a) = 0$ if $a \notin NE^*$. Then, note that for any player $i$, any $p \in [0,1]$ and any $a_i \neq a_i^i, a_i^j$, we have:

$$p \cdot u_i^*(a_i^i, a_j^j) + (1-p) \cdot u_i^*(a^i) > p \cdot u_i^*(a_i, a_j^j) + (1-p) \cdot u_i^*(a_i, a_j^*(a_i)),$$

because $u_i^*(a^i) > u_i^*(a_i, a_j^*(a_i))$ for any $a_i \neq a_i^i$ by definition, and $u_i^*(a_i^i, a_j^j) \geq u_i^*(a_i, a_j^j) = 0$ for any $a_i \neq a_i^j$. Hence, $a_i^i$ dominates all $a_i \neq a_i^j, a_i^i$ for any $p$, and it is better than $a_j^i$ for high $p$, and worse than $a_i^j$ for low $p$. It follows that $\mathcal{B}_i^2 = \{a_i^i, a_i^j\}$. But then, at the next round, for any $p, q \in [0,1]$ and any $a_i \neq a_i^i, a_i^j$ we have:

$$\begin{aligned} pq \cdot u_i^*(a_i^i, a_j^j) + p(1-q) \cdot u_i^*(a^i) + (1-p) \cdot u_i^*(a^i) \\ > pq \cdot u_i^*(a_i, a_j^j) + p(1-q) \cdot u_i^*(a^i) + (1-p) \cdot u_i^*(a_i, a_j^*(a_i)), \end{aligned}$$

and by the same argument as before only $a_i^i$ and $a_i^j$ can be a best response for some $p$ and $q$. It follows that $\mathcal{B}_i = \{a_i^i, a_i^j\} \subseteq R_i^*$, and hence $RP_i \subseteq \{a_i^i, a_i^j\}$. The result then follows from Theorem 1. ■

**Proof of Lemma 3.** Let $\bar{u}_i := u_i^*\,(a)$ for all $a \notin NE^*$. Then, for any $a^* \in NE^*$, by definition it must be that $u_i^*(a^*) > \bar{u}_i$ for all $i$. ■

**Proof of Lemma 4.** By definition, in any game which satisfies Assumption 1, $u_i^*(a^i) > u_i^*(a_i, a_j^*(a_i)$ for all $a_i \neq a_i^i$. Furthermore, in any game, any $\hat{a} \in NE^*$ can be written as $\hat{a} = (\hat{a}_i, a_j^*(\hat{a}_i))$. Hence, if $\hat{a} \in NE^*$, it must be that $u_i^*(a^i) \geq u_i^*(\hat{a})$. But by Definitions 1, if $G^*$ is an Agreement game this is only possible if $a^i \in NE^*$.

If $G^*$ is a Coordination game, letting $A_i = \{a_i(1), \ldots, a_i(n^*)\}$ denote set of actions, ordered such that all profiles on the main diagonal, of the form $(a_i(n), a_j(n))$, are a Nash equilibrium, we have that for every $i$, $j \neq i$ and every $n = 1, \ldots, n^*$, $a_j^*(a_i(n)) = a_j(n)$. Hence, suppose that $a^i \notin NE^*$, and let $a_i(n) := a_i^i$ and $a_j(n) := a_j^*(a_i^i)$. Then, either $u_j^*(a_j', a_i(n)) > u_j^*(a_j(n), a_i(n))$ for some $a_j'$, which contradicts that $(a_j(n), a_i(n)) \in NE^*$, or $u_i^*(a_i', a_j(n)) > u_i^*(a_i(n), a_j(n))$ for some $a_i'$, which contradicts $a_i^*(a_j(n)) = a_i(n)$. Hence, if $G^*$ is a coordination game, $a^i \in NE^*$. ■

**Proof of Proposition 2.** For part $(i)$, since $G^*$ is a zero-sum game, if $a^* \in NE^*$ it follows from Assumption 1 that for each $i$, $\arg\max_{a_i \in A_i} \min_{a_j \in A_j} u_i^*(a_i, a_j) = \{a_i^*\}$. In a zero-sum game, for any $a_i \in A_i$ we have $|\arg\min_{a_j \in A_j} u_i^*(a_i, a_j) = \arg\max_{a_j \in A_j} u_j(a_i, a_j)| = 1$; let $s_j^*(a_i)$ denote such unique element. Thus, $\arg\max_{a_i \in A_i} \min_{a_j \in A_j} u_i^*(a_i, a_j) = \{a_i^*\}$ if and only if $\arg\max_{a_i \in A_i} u_i^*(a_i, s_j^*(a_i)) = \{a_i^*\}$, that is if and only if $a_i^* = a_i^i$. Hence, $a^* = (a_i^i, a_j^j)$. However, since $a^*$ is an equilibrium, it also satisfies $a_i^j = a_i^i$ for all $i$. Hence, $a^* = a^1 = a^2$. But then, it is easy to verify that $\mathcal{B}^2 = \{a^*\}$, and since $a_i^*$ is a strict best response to $a_j^*$ in both the static and dynamic case, no further actions are added at the subsequent iterations: $\mathcal{B}^k = \mathcal{B}^2$ for all $k \geq 2$. Hence, $\mathcal{B} = \{a^*\}$. Part $(ii)$ follows from part $(i)$ and Theorem 1. ■

**Proof of Proposition 3.** Fix player $i$. We know from Theorem 1 that there exists some dense subset $\check{T}_i \subseteq T_i^*$ such that $|R_i(t_i)| = 1$ and $R_i(t_i) = RP_i(t_i)$ for any $t_i \in \check{T}_i$. Notice now that since $G^*$ is a common interest game, we know that $a_i^i = a_i^j = a_i^*$. Thus, it follows from Assumption 1 that $a_i^j = a_j^i$, and hence, that $\mathcal{B}_i = \{a_i^*\}$. Then, the fact that $RP_i$ always yields a subset of $\mathcal{B}_i$ implies that $R_i(t_i) = RP_i(t_i) = \{a_i^*\}$ for any $t_i \in \check{T}_i$. Now, note that it follows from upper-hemicontinuity of $R_i$ that $T_i' := \{t_i \in T_i^* : R_i(t_i) = \{a_i^*\}\}$ is open, and clearly, we have $\check{T}_i \subseteq T_i'$. Thus, $T_i'$ is and open and dense subset of $T_i^*$ satisfying that $R_i(t_i) = \{a_i^*\}$ for every $t_i$ in it. ■

**Proof of Proposition 4.** To see part $(i)$ notice that if $RP_j(t_j^{CB(\omega^0)}) = \{a_j\}$, then, for any player $i$'s conjecture $\mu_i \in \bigcap_{k \geq 0} \Delta(RP_j^k) \cap C_i(t_i^{CB(\omega^0)})$, $\mu_i[(t_j^{CB(\omega^0)}, a_j)] = 1$. Since $RP_i(t_i^{CB(\omega^0)}) = \{a_i\}$, it follows that $a_i$ is a best reply to $a_j$. Hence, $a$ is a Nash equilibrium of baseline game $G^*$.

For part $(ii)$, notice first that we know from part $(i)$ of Theorem 1 that $RP(t) = \{a\}$ for any $t \in \Lambda^k(t^{CB(\omega^0)})$. We also know from part $(ii)$ of Theorem 1 that $RP$ is upper-hemicontinuous, and thus, it follows that there exists some open set $V$ such that $\Lambda^k(t^{CB(\omega^0)}) \subseteq V$ and $RP(t) = \{a\}$ for any $t \in V$. Now, we know from part $(iii)$ of Theorem 1 that there exists some open and dense set $W$ such that $R(t) = RP(t)$ for any $t \in W$. It follows that $U := V \cap W$ is an open set such that $(a)$ $R(t) = \{a\}$ for any $t \in U$, and $(b)$ $\Lambda^k(t^{CB(\omega^0)})$ is a subset of the topological closure of $U$. ■

# E Asynchronous Moves

To avoid ambiguity, in the present section we denote by $R_i^e$, $RP_i^e$ and $RP_i^{e,\mathcal{A}}$ correspondences $R_i$, $RP_i$ and $RP_i^{\mathcal{A}}$, respectively, when they are defined in domain $T_i^e$. First, we check in the following lemma that whether agreement on asynchronicity is modeled via knowledge or via belief turns out to be immaterial in terms of behavior:

**Lemma 13** *For any player $i$, any $k \geq 0$ and any $t_i \in T_i^e$ the following two hold:*

*(i)* $R_i^{k,e}(t_i) = R_i^k(\phi_i^e(t_i))$.

*(ii)* $RP_i^{k,e,\mathcal{A}}(t_i) = RP_i^{k,\mathcal{A}}(\phi_i^e(t_i))$.

**Proof.** For any player $i$ let map $\Phi_i^e : \{\mu_i \in C_i(t_i) : t_i \in T_i^e\} \to \{\mu_i' \in C_i(t_i) : t_i \in T_i^*(e)\}$ be given by,

$$\Phi_i^e(\mu_i)[E] := \mu_i\left[\left\{(t_j, s_j) \in T_j^e \times S_j : (\phi_j^e(t_j), s_j) \in E\right\}\right].$$

for any $\mu_i \in \{\mu_i \in C_i(t_i) : t_i \in T_i^e\}$ and any measurable $E \subseteq T_j^e \times S_j$. We saw in Lemma 5 that $\phi_i^e$ is a homeomorphism between $T_i^e$ and $T_i^*(e)$ satisfying that $\tau_i^e(t_i)[E] = \tau_i^e(\phi_i^e(t_i))[\phi_j^e(E)]$ for any type $t_i \in T_i^e$ and any measurable $E \subseteq T_j^e$. It follows immediately that $\Phi_i^e$ is a well-defined homeomorphism that satisfies a similar belief-preserving property, namely that for any type $t_i \in T_i^e$ and any $\mu_i \in C_i(t_i)$,

$$\mu_i[E] = \Phi_i^e(\mu_i)\left[\left\{(\phi_j^e(t_j), s_j) \in T_j^* \times S_j : (t_j, s_j) \in E\right\}\right],$$

for any measurable $E \subseteq T_j^e \times S_j$. In particular, this implies that:

$(a)$ For any $t_i \in T_i^e$ and any $\mu_i C_i(t_i)$ we have that $BR_i(\mu_i; t_i) = BR_i(\Phi_i^e(\mu_i); t_i)$ and that for any $k \geq 0$,

$$\mu_i\left[R_j^{k,e}\right] = \Phi_i^e(\mu_i)\left[\left\{(\phi_j^e(t_j), s_j) \in T_j^* \times S_j : (t_j, s_j) \in R_j^{k,e}\right\}\right],$$
$$\mu_i\left[RP_j^{k,e,\mathcal{A}}\right] = \Phi_i^e(\mu_i)\left[\left\{(\phi_j^e(t_j), s_j) \in T_j^* \times S_j : (t_j, s_j) \in RP_j^{k,e,\mathcal{A}}\right\}\right].$$

$(b)$ For any $t_i \in T_i^*(e)$ and any $\mu_i C_i(t_i)$ we have that $BR_i(\mu_i; t_i) = BR_i(\Phi_i^{e,-1}(\mu_i); t_i)$ and that for any $k \geq 0$,[33]

$$\mu_i\left[R_j^k\right] = \Phi_i^{e,-1}(\mu_i')\left[\left\{(t_j, s_j) \in T_j^e \times S_j : (\phi_j^e(t_j), s_j) \in R_j^k\right\}\right],$$
$$\mu_i\left[RP_j^{k,\mathcal{A}}\right] = \Phi_i^{e,-1}(\mu_i')\left[\left\{(t_j, s_j) \in T_j^* \times S_j : (\phi_j^e(t_j), s_j) \in RP_j^{k,\mathcal{A}}\right\}\right].$$

Relying in these equivalences it is immediate to check claims $(i)$ and $(ii)$ by induction on $k$ starting from the obvious fact that $(i)$ and $(ii)$ are trivially true for $k = 0$. ■

In addition, for the proof of Theorem 2 we require the following technical remark:

**Lemma 14** *For any player $i$, $RP_i^{\mathcal{A}}$ is a refinement of $RP_i$.*

**Proof.** We check by induction on $k$ that $RP_i^{k,\mathcal{A}}(t_i) \subseteq RP_i^k(t_i)$ for any $k \geq 0$ and and $t_i \in T_i^*$. The initial step ($k = 0$) holds trivially so let's focus on the inductive one. Suppose

[33]Remember that $\Phi_i^e$ is bijective; we denote its inverse map by $\Phi_i^{e,-1}$,

that $k \geq 0$ is such that the claim holds. We verify then that it also does for $k+1$. Fix player $i$, type $t_i \in T_i^*$, strategy $s_i \in RP_i^{k+1,\mathcal{A}}(t_i)$ and conjecture $\mu_i \in \Delta(RP_j^{k,\mathcal{A}}) \cap C_i(t_i)$ such that $s_i \in BR_i(\mu_i; t_i)$. Since, by the induction hypothesis, $RP_j^{k,\mathcal{A}}$ is a refinement of $RP_j^k$, then we know that $\Delta(RP_j^{k,\mathcal{A}}) \subseteq \Delta(RP_j^k)$, and thus, it follows that $\mu_i \in \Delta(RP_j^k) \cap C_i(t_i)$. Hence, $s_i \in RP_i^{k+1}(t_i)$. ■

We can now proceed to the proof of Theorem 2:

**Proof of Theorem 2.** Non-emptiness and upper-hemicontinuity follow immediately from Lemmas 2 and 5 and part $(ii)$ of Theorem 1. To see part $(iii)$, fix finite type $t_i \in T_i^e$ and strategy $s_i \in RP_i^{e,\mathcal{A}}(t_i)$. Then, we know from Lemma 5 that there exists some $t_i^* \in T_i^*(e)$ such that $\phi_i^e(t_i) = t_i^*$, and from Lemma 13, that $RP_i^{e,\mathcal{A}}(t_i) = RP_i^{\mathcal{A}}(t_i^*)$, and therefore, that $s_i \in RP_i^{\mathcal{A}}(t_i^*) \subseteq RP_i(t_i^*)$. Let's proceed in three steps:

$(a)$ Notice first that for any $a_i \in \mathcal{A}_i$, dominance type $t_i^{a_i}$ constructed similarly as in Lemma 10 is an element of $T_i^*(e)$. To see it, for $a_i \in \mathcal{A}_i^1 = \{a_i^e\}$, set $t_i^{a_i} = t_i^{CB(\omega^e)}$ if $i = e$, and $t_i^{a_i} = t_i^{CB(\omega^0)}$ otherwise; clearly, $SP_i(t_i^{a_i}) = \{a_i\}$. For $a_i \in \mathcal{A}_i \setminus \{a_i^e\}$, construct $t_i^{a_i}$ as done in Lemma 10, but initializing the construction as done for the case of $a_i^e$. Since for each $k \geq 1$ and each $a_i \in \mathcal{A}_i^k$ type $t_i^{a_i}$ only assigns positive probability to types $t_j^{a_j}$ where $a_j \in \mathcal{A}_j^{k-1}$, it is easy to verify by induction that $t_i^{a_i} \in T_i^*(e)$.

$(b)$ Next, proceed as in Lemma 11 and pick type space $\mathcal{T}$ that contains a type $t_i$ such that $\phi_i^e(t_i) = t_i$. Then, define for each $\nu \in \mathbb{N}$ type space $\mathcal{T}^\nu$ as done in Lemma 11, by using the dominance types obtained in the previous bullet. Since every type in each $\mathcal{T}^\nu$ is a combination of a type in $\mathcal{T}$ and a dominance solvable type as the ones in the previous bullet, clearly, $\phi_i(T_i^\nu) \subseteq T_i^*(e)$. Then, it follows from Lemma 11 that for each $\nu \in \mathbb{N}$ we can pick $\hat{t}_i^\nu \in \phi_i(T_i^\nu) \cap \phi_i^e(T_i^e)$ such that $s_i \in SP_i(\hat{t}_i^\nu)$, and that, furthermore, $(\hat{t}_i^\nu)_{\nu \in \mathbb{N}}$ converges to $t_i^*$.

$(c)$ Now, following Lemma 12, for each $\nu \in \mathbb{N}$, pick sequence $(\hat{t}_i^{\nu,k})_{k \in \mathbb{N}}$ converging to $\hat{t}_i^\nu$ and such that $R_i(\hat{t}_i^{\nu,k}) = RP_i(\hat{t}_i^{k,\nu}) = \{s_i\}$ for any $k \in \mathbb{N}$. Now, note first that each $\hat{t}_i^{1,\nu}$ is a dominance solvable type as the ones defined in bullet $(i)$, being thus in $T_i^*(e)$, and second, that each $t_i^{k,\nu}$ only assigns positive probability to types $t_j^{k-1,\nu}$ of player $j$ (that are obtained from pairs $(t_j, a_j)$ in the support of some conjecture $\mu_i \in C_i(\hat{t}_i^\nu) \cap \Delta(SP_i)$ for which $BR_i(\mu_i; \hat{t}_i^\nu) = \{s_i\}$). Hence, similarly as done in bullet $(a)$, it is easily verifiable by induction that $\hat{t}_i^{k,\nu} \in T_i^*(e)$ for any $k, \nu \in \mathbb{N}$.

Then, let $(t_i^\nu)_{\nu \in \mathbb{N}}$ be defined by picking, for each $\nu \in \mathbb{N}$, the unique $t_i^\nu \in \phi_i^{e,-1}(\hat{t}_i^{\nu,\nu})$. It follows from Lemma 13 that $(t_i^\nu)_{\nu \in \mathbb{N}}$ is a sequence in $T_i^e$ that converges to $t_i$ and such that $R_i^e(t_i^\nu) = RP_i^{e,\mathcal{A}}(t_i^\nu) = \{s_i\}$ for any $\nu \in \mathbb{N}$. ■

Propositions 5 and 7 are proved by adapting the proofs of Propositions 3 and 4, respectively, in the obvious way. We thus focus on the proof of Proposition 6.

**Proof of Proposition 6.** We know from Lemma 6 that $R_e(t_e^{CB(\omega^e)}) = \{a_e^e\}$. It follows immediately that $R_\ell(t_\ell^{CB(\omega^e)}) = \{a_\ell^e\}$. Hence, we know by upper-hemicontinuity of $R$ that there exists some open neighborhood of type profile $t^{CB(\omega^e)}$, $U$, such that $R(t'^e\} \, \forall t' \in U$. ■

# References


1. Alaoui, L. and A. Penta (2016) "Endogenous Depth of Reasoning," *Review of Economic Studies*, 83, 1297–1333.

2. Ambrosio L., N. Gigli. and G. Savaré (2005) *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, Lectures in Mathematics, ETH Zürich, Birkhäuser.

3. Ambrus, A. and Y. Ishii (2015) "On Asynchronicity of Moves and Coordination," *ERID Working Paper* Number 185.

4. Amershi, A.H., A. Sadanand and V. Sadanand (1989) "Manipulated Nash Equilibria," *mimeo*, University of Guelph,

5. Amershi, A.H., A. Sadanand and V. Sadanand (1992) "Player importance and Forward Induction," *Economic Letters*, 38, 291–297.

6. Artemov, G., T. Kunimoto and R. Serrano (2013) "Robust Virtual Implementation with Incomplete Information: Towards a Reinterpretation of the Wilson Doctrine," *Journal of Economic Theory*, 148, 424–447.

7. Aumann, R. and S. Sorin (1999) "Cooperation and bounded recall," *Games and Economic Behavior*, 1, 5–39.

8. Bagwell, K. (1995) "Commitment and Observability in Games," *Games and Economic Behavior*, 8, 271–280.

9. Battigalli, P., A. Di Tillio, E. Grillo and A. Penta (2011) "Interactive Epistemology and Solution Concepts for Games with Asymmetric Information," the *B.E. Journal in Theoretical Economics: Advances*, 11, article 6.

10. Ben-Porath, E. (1997) "Rationality, Nash Equilibrium and Backwards Induction in Perfect Information Games," *Review of Economic Studies*, 64, 23–46.

11. Bergemann, D. and S. Morris (2005) "Robust Mechanism Design," *Econometrica*, 73, 1521–1534.

12. Bergemann, D. and S. Morris (2009) "Robust Implementation in Direct Mechanisms," *Review of Economic Studies*, 76, 1175–1204.

13. Bergemann, D. and S. Morris (2013) "Robust Predictions in Games With Incomplete Information," *Econometrica*, 81, 1251–1308.

14. Bergemann, D. and S. Morris (2016) "Bayes correlated equilibrium and the comparison of information structures in games," *Theoretical Economics*, 11, 487–522.

15. Bernheim, D. (1984) "Rationalizable Strategic Behavior," *Econometrica* 52, 1007–1028.

16. Binmore, K. (1987-88) “Modeling Rational Players,” *Economics and Philosophy*, 3:179–214; 4:9–55.

17. Bonanno, G. (1992) “Deterrence, Observability and Awareness,” *Economic Notes*, 21, 307–315.

18. Calcagno, R., Y. Kamada, S. Lovo and T. Sugaya (2014) “Asynchronicity and coordination in common and opposing interest games,” *Theoretical Economics*, 3:179–214; 4:9–55.

19. Camerer, C. (2003) *Behavioral Game Theory*, Princeton, NJ, Princeton Univ. Press.

20. Carlsson, H. and E. van Damme (1993) “Global Games and Equilibrium Selection,” *Econometrica,* 61, 989–1018.

21. Chen, Y., S. Takahashi and S. Xiong (2014) “The Weinstein-Yildiz Critique and Robust Predictions with Arbitrary Payoff Uncertainty,” *mimeo.*

22. Cooper, R., D. DeJong, R. Forsythe and T. Ross (1993) “Forward Induction in the Battle of the Sexes Games,” *American Economic Review*, 83, 1303–1316.

23. Costa-Gomes, M.A. and V.P. Crawford (2006) “Cognition and Behavior in Two-Person Guessing Games: An Experimental Study,” *American Economic Review,* 96, 1737–1768.

24. Costa-Gomes, M.A, V.P. Crawford, and B. Broseta. (2001) “Cognition and Behavior in Normal-Form Games: An Experimental Study,” *Econometrica*, 69, 1193–1235.

25. Crawford, V.P., M.A. Costa-Gomes, and N. Iriberri. (2013) “Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications,” *Journal of Economic Literature*, 51, 5–62.

26. Crawford, V.P. and N. Iriberri (2007) “Level-$k$ Auctions: Can a Nonequilibrium Model of Strategic Thinking Explain the Winner’s Curse and Overbidding in Private-Value Auctions?,” *Econometrica*, 75, 1721–1770.

27. Dasgupta, A., Steiner, J. and C. Stewart (2010) “Dynamic Coordination with Individual Learning,” *Games and Economic Behavior*, 74, 83–101.

28. van Damme, E. and S. Hurkens (1996) “Commitment Robust Equilibria and Endogenous Timing,” *Games and Economic Behavior*, 15, 290–311.

29. van Damme, E. and S. Hurkens (1997) “Games with Imperfectly Observable Commitment,” *Games and Economic Behavior*, 21, 282–308.

30. van Damme, E. and S. Hurkens (1999) “Endogenous Stackelberg Leadership,” *Games and Economic Behavior*, 28, 105–129.

31. van Damme, E. and S. Hurkens (2004) “Endogenous Price Leadership,” *Games and Economic Behavior*, 47, 404–420.

32. Dekel, E. and D. Fudenberg (1990) “Rational Behavior With Payoff Uncertainty,” *Journal of Economic Theory*, 52, 243–267.

33. Dekel, E., D. Fudenberg and S. Morris (2006) “Topologies on Types,” *Theoretical Economics*, 1, 275–309.

34. Dekel, E., D. Fudenberg and S. Morris (2007) “Interim Correlated Rationalizability,” *Theoretical Economics*, 2, 15–40.

35. Ely, J. and L. Doval (2016) “Sequential Information Design,” *mimeo.*

36. Ely, J. and Pęski, M. (2006) “Hierarchies of belief and interim rationalizability,” *Theoretical Economics*, 1, 19–65.

37. Frankel, D.M., S. Morris and A. Pauzner (2003) “Equilibrium Selection in Global Games with Strategic Complementarities,” *Journal of Economic Theory*, 108, 1–44.

38. Fudenberg, D. and D.Levine (1998), *The Theory of Learning in Games*, Cambridge, MA, MIT Press.

39. Guesnerie, R. (2005) *Assessing Rational Expectations 2: “Eductive Stability” in Economics*, Cambridge, MA, MIT Press.

40. Haywood, O.G. Jr. (1954) “Military decision and game theory,” *Journal of Operations Research Society of America*, 2, 365–385.

41. Harsanyi, J.C. (1981) “Solutions for Some Bargaining Games Under the Harsanyi-Selten Solution Theory, Part II: Analysis of Specific Bargaining Games,” *CP-432 Center for Research in Management Science*, University of California, Berkeley.

42. Iriberri, N. and V. Crawford (2007), “Fatal Attraction: Salience, Naïveté, and Sophistication in Experimental “Hide-and-Seek” Games,” *American Economic Review*, 97, 1731–1750.

43. Lagunoff, R. and A. Matsui (1997) “Asynchronous Choice in Repeated Coordination Games,” *Econometrica* 65, 1467–1477.

44. Kajii, A. and S. Morris (1997) “The Robustness of Equilibria to Incomplete Information,” *Econometrica*, 65, 1283–1309.

45. Kalai, E. and D. Samet (1984) “Persistent Equilibria in Strategic Games,” *International Journal of Game Theory*, 13, 129–144.

46. Kamada, Y. and M. Kandori (2017a) “Revision Games. Part I: Theory,” *mimeo.*

47. Kamada, Y. and M. Kandori (2017b) "Revision Games. Part II: Applications," *mimeo.*

48. Kovac, E. and J. Steiner (2013) "Reversibility in Dynamic Coordination Problems," *Games and Economic Behavior*, 77, 298–320.

49. Kreps, D. (1990), *Game Theory and Economic Modeling*, Oxford, UK, Oxford Univ. Press.

50. Li, H. and M. Peters (2017) "Unobserved Mechanisms," *mimeo.*

51. Luce, R.D. and H. Raiffa (1957), *Games and Decisions*, Wiley and Sons, New York, NY, USA.

52. Makris, M. and L. Renou (2018) "Information design in multi-stage games," *mimeo*

53. Mathevet, L. and J. Steiner (2013) "Tractable Dynamic Global Games and Applications," *Journal of Economic Theory*, 148, 2583–2619.

54. Mertens, J.F. and S. Zamir (1985) "Formulation of Bayesian Analysis for Games with Incomplete Information," *International Journal of Game Theory*, 14, 1–29.

55. Morris, S. (2000) "Contagion", *Review of Economic Studies*, 67, 57–78.

56. Morris, S. and H.S. Shin (1998) "Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks,"*American Economic Review*, 88, 587–597

57. Myerson, R. (1991) *Game Theory: The Analysis of Conflict*, Harvard University Press.

58. Nagel, R. (1995) "Unraveling in Guessing Games: An Experimental Study," *American Economic Review*, 85, 1313–1326.

59. Von Neumann, J. and O. Morgenstern (1944), *The Theory of Games and Economic Behavior*, Princeton University Press, Princeton, NJ, USA.

60. Von Neumann, J. and O. Morgenstern (1947), *The Theory of Games and Economic Behavior, 2nd rev. ed.*, Princeton University Press, Princeton, NJ, USA.

61. Ollar, M. and A. Penta (2017) "Full Implementation and Belief Restrictions," *American Economic Review*, 108, 2243–2277.

62. Oury, M. and O. Tercieux (2012) "Continuous Implementation," *Econometrica*, 80, 1605–1637.

63. Pearce, D. (1984) "Rationalizable Strategic Behavior and the Problem of Perfection,"*Econometrica*, 52, 1029–1050.

64. Penta, A. (2012) "Higher Order Uncertainty and Information: Static and Dynamic Games," *Econometrica*, 80, 631–660.

65. Penta, A. (2013) "On the Structure of Rationalizability on Arbitrary Spaces of Uncertainty," *Theoretical Economics*, 8 (2013), 405-430

66. Penta, A. (2015) "Robust Dynamic Implementation," *Journal of Economic Theory*, 160, 280–316

67. Penta, A. and P. Zuazo-Garin (2017) "Fear of Leaks and Strategic Uncertainty in Games of Conflict," *mimeo.*

68. Peters, M. (2015) "Can Mechanism Designers Exploit Buyers' Market Information," *mimeo.*

69. Rubinstein, A. (1989) "The Electronic Mail Game: Strategic Behavior Under 'Almost Common Knowledge'," *American Economic Review*, 79, 385–391.

70. Salcedo, B. (2017) "Interdependent Choices," *mimeo.*

71. Samuelson, L. (1998) *Evolutionary Games and Equilibrium Selection*, Cambridge, MA, MIT Press.

72. Schelling, T.C. (1960) *The Strategy of Conflict*, Harvard University Press.

73. Stahl, D. and P. Wilson (1994) "Experimental Evidence on Players' Models of Other Players," *Journal of Economic Behavior & Organization*, 25, 309–327.

74. Stahl, D. and P. Wilson (1995) "On Players' Models of Other Players: Theory and Experimental Evidence," *Games and Economic Behavior*, 10, 218–254.

75. Steiner, J. and C. Stewart (2008) "Contagion through Learning," *Theoretical Economics*, 3, 431–458.

76. Solan, E. and L. Yariv (2004) "Games with Spionage," *Games and Economic Behavior*, 47, 172–199.

77. Sugden, R. (1995) "A Theory of Focal Points," *Economic Journal*, 105, 533–550.

78. Weinstein, J. and M. Yildiz (2007) "A Structure Theorem for Rationalizability with Application to Robust Predictions of Refinements," *Econometrica,* 75, 365–400.

79. Weinstein, J. and M. Yildiz (2011) "Sensitivity of Equilibrium Behavior to Higher Order Beliefs in Nice Games," *Games and Economic Behavior*, 72, 288–300.

80. Weinstein, J. and M. Yildiz (2013) "Robust Predictions in Infinite-horizon Games– An Unrefinable Folk Theorem," *Review of Economic Studies*, 80, 365–394.

81. Weinstein, J. and M. Yildiz (2016) "Reputation without Commitment in Finitely-Repeated Games" *Theoretical Economics*, 11, 157–185.

82. Zuazo-Garin, P. (2017), "Uncertain Information Structures and Backward induction," *Journal of Mathematical Economics*, 71, 135–150.