Rationalizability, Observability and Common Knowledge

Antonio Penta
Peio Zuazo-Garin

July 2019

Barcelona GSE Working Paper Series

Working Paper nº 1106

Rationalizability, Observability and Common Knowledge*

Antonio Penta† Peio Zuazo-Garin‡

July 1, 2019

Abstract

We study the strategic impact of players' higher order uncertainty over the observability of actions in general two-player games. More specifically, we consider the space of all belief hierarchies generated by the uncertainty over whether the game will be played as a static game or with perfect information. Over this space, we characterize the correspondence of a solution concept which represents the behavioral implications of Rationality and Common Belief in Rationality (RCBR), where 'rationality' is understood as *sequential* whenever a player moves second. We show that such a correspondence is generically single-valued, and that its structure supports a robust refinement of rationalizability, which often has very sharp implications. For instance: (i) in a class of games which includes both zero-sum games with a pure equilibrium and coordination games with a unique efficient equilibrium, RCBR generically ensures efficient equilibrium outcomes; (ii) in a class of games which also includes other well-known families of coordination games, RCBR generically selects components of the Stackelberg profiles; (iii) if common knowledge is maintained that player 2's action is *not* observable (e.g., because 1 is commonly known to move earlier, etc.), in a class of games which includes of all the above RCBR generically selects the equilibrium of the static game most favorable to player 1.

Keywords: eductive coordination – extensive form uncertainty – first-mover advantage – Krpes hypothesis – higher order beliefs – rationalizability – robustness – Stackelberg selections

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 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

*We thank Pierpaolo Battigalli, Eddie Dekel, Laura Doval, Glenn Ellison, Drew Fudenberg, Lones Smith, Stephen Morris, Bill Sandholm, Joel Sobel, Jakub Steiner, Joel Watson and Muhamet Yildiz for helpful comments and suggestions. We thank seminar audiences at the Univ. Pública de Navarra, Cambridge, UW-Madison, UCSD, UPV/EHU, PSE, CERGE-EI , UPF, TSE, Groningen, Bonn, EUI , and participants of the 2017 SISL Summer Conference (Caltech), 2017 HKU-UBC Summer Workshop in Economic Theory (Hong-Kong Univ.), 2017-StratEmon Workshop (Bocconi), 2018 Southampton-Bristol Economic Theory Workshop (Southampton), 2018 Warwick Economic Theory Workshop (Warwick Univ.), 2018 LOFT Conference (Bocconi). Antonio Penta acknowledges the financial support of the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2015-0563) and of the ERC Starting Grant #759424.

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

‡Universitat Pompeu Fabra. E-mail: peio.zuazo@upf.edu

form has hardly been challenged.[1] Yet, uncertainty over the extensive form is key to many strategic situations, and in many economic settings it need not match exactly the kind of common knowledge assumptions which are implicit in standard models. The reliability of such models therefore depends on whether the predictions they generate are robust to this kind of model misspecification.

For instance, when we study firms interacting in a market, we often model the situation as a static game (Cournot competition, simultaneous entry, technology adoption, etc.), or as a dynamic one (e.g., Stackelberg, sequential entry, sequential technology adoption, etc.). But, in the former case, this not only presumes that firms' decisions are made without observing other firms' choices, 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 observability of actions in ways which would be impossible to model with absolute precision. 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 observability of actions were misspecified in our model.

To address this question, we consider the space of all belief hierarchies generated by players' uncertainty over whether a two-player game will be played as a static game, i.e. with no information about others' moves, or sequentially, with perfect information. Over this space, we characterize the correspondence of a solution concept – formally denoted by R – which represents the behavioral implications of Rationality and Common Belief in Rationality (RCBR), where the term 'rationality' is understood as *sequential*, whenever the game is dynamic.[2] For general two-player games, we show that R is generically single-valued, and that it admits a robust and non-empty refinement which characterizes the *regular predictions* of RCBR, i.e. those which do not depend on knife-edge, non-generic restrictions on the belief hierarchies. We then explore the implications of these results in classes of games in which the they are especially sharp or significant.

For example, we show that in a class of games which includes common interest games (Aumann and Sorin (1989)), coordination games with a unique efficient equilibrium (e.g., stag-hunt, pure coordination, etc.), but also zero-sum games with a pure equilibrium, RCBR generically selects the efficient equilibrium actions. Aside from the sharpness of the refinement it supports for these games, this result shows that higher order uncertainty over the extensive form may serve as a mechanism for equilibrium coordination based on purely introspective reasoning. This is especially significant because the possibility that correct conjectures can

[1] Some papers have studied commonly known structures to represent players' uncertainty over the extensive form (most notably, Robson (1994), Reny and Robson (2004) and Kalai (2004), etc.), 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. We discuss the related literature in Section 7.

[2] Under a genericity assumption on payoffs, the behavioral implications of RCBR in our setting are conveniently obtained applying iterated strict dominance to the interim normal form of the game with extensive form uncertainty, preceded by one round of weak dominance only for those types who observe the opponent's action – the round of weak dominance serves to capture *sequential* rationality. R is thus a hybrid of Interim Correlated Rationalizability (Dekel et al., 2007) and Dekel and Fudenberg's (1990) $S^{\infty}W$ procedure, and is weaker than virtually any standard solution concept based on sequential rationality.

be achieved on the basis of purely 'eductive' mechanisms (Binmore (1987-88)), in the absence of focal points and without any information on past interactions, is generally met with skepticism.[3] Our result shows that, in the presence of higher order uncertainty over the observability of actions, equilibrium coordination emerges endogenously as the generic implication of standard assumptions of RCBR, without appealing to external, non-mathematical properties of the game nor to notions of bounded rationality. For zero-sum games with a pure equilibrium, this result also implies that, for a generic set of belief hierarchies, the maxmin solution coincides with the unique implication of RCBR , thereby 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 a class of games which includes all of the above, as well as other well-known families of coordination games (e.g., Harsanyi (1981) and Kalai and Samet's (1984) 'unanimity' games), we find that for a generic set of belief hierarchies, RCBR implies that players choose components of the Stackelberg profiles, regardless of the actual observability of actions.

We also characterize the robust predictions in environments with 'one-sided' uncertainty, 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. Such one-sided uncertainty arises naturally in a number of settings, for instance when moves are chosen at different points in time, with a commonly known order. But it is also relevant in any situation in which players commonly agree that only one of them is committed to ignoring the other's action, or that only the actions of one player are effectively irreversible, etc. In these settings, the analysis delivers particularly striking results: In a class of games which encompasses as special cases all of those discussed above, we show that RCBR generically selects the equilibrium of the static game which is most favorable to the earlier mover (or, more generally, the player who is commonly known to *not* observe the opponent's move). Hence, a first-mover advantage is *pervasive* in these games: it arises for a generic set of types, regardless of whether the action is actually observable, including for types who share arbitrarily many (but finite) orders of mutual belief that the action is *not* observable.

This result has important strategic implications, in that it points at the impact that mechanisms to establish common knowledge of one-sided uncertainty may have in the presence of higher order uncertainty over the observability of actions. As discussed, various kinds of mechanisms may produce this kind of uncertainty, but perhaps the simplest and most obvious to consider is the one associated to a commonly known order of moves: Within this context, our result suggests that, by determining the direction of the one-sided uncertainty, a commonly known *timing* of moves (plus irreversibility of choices) may determine the attribution of the strategic advantage, independent of the actual observability of actions. 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 wis-

[3]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. Questions of eductive stability have been pursued in economics both in partial and general equilibrium settings (see, e.g., Guesnerie (2005) and references therein).

dom. For instance, asynchronous moves in the Battle of the Sexes systematically select the Nash equilibrium most favorable to the first mover (see Camerer (2003), Ch.7, and references therein), 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 the unique regular prediction consistent with RCBR, when one considers higher order uncertainty over the observability of actions.[4]

The discussion above suggests that perturbing common knowledge assumptions on the observability of actions has very different implications from those on payoffs. Weinstein and Yildiz (2007,WY), in particular, show that when the space of payoff uncertainty is 'rich', the only predictions which are robust to even small mispecifications of the belief hierarchies are those which can be made based on rationalizability alone. But our general results also have important similarities with WY's. More specifically, similar to WY, we show that R is everywhere upper hemicontinuous (u.h.c.) and is generically single-valued. But while WY's result also implies that no refinement of rationalizability is u.h.c., we show that there exists a non-empty and u.h.c. refinement of R, which we call RP, with the property that whenever it delivers multiple predictions, any such predictions is uniquely selected by both R and RP for some arbitrarily close hierarchies of beliefs. Moreover, such nearby uniqueness result only holds for the predictions included in RP. Hence, it turns out that not only is RP (i) the *strongest robust refinement* consistent with RCBR; but it also (ii) coincides with the RCBR predictions *generically* on the universal type space; and (iii) it characterizes, *everywhere* on the universal type space, the predictions of RCBR which are *regular* in the sense of not depending on knife-edge situations, associated to non-generic subsets of belief hierarchies.

Our structure theorem thus describes a very different correspondence from WY's. In both cases, multiplicity is only possible within non generic sets of belief hierarchies. But while in WY, when multiplicity occurs, it cannot be robustly refined away, because any of the rationalizable outcomes is uniquely selected in an open set of arbitrarily close types, in our space of uncertainty there may be actions (specifically, those in R but not in RP) which are rationalizable *only* within non-generic sets of belief hierarchies. No analogous of this

[4]This is not to say that the logic of our results necessarily provides a behaviorally accurate model of strategic thinking (see, e.g., Crawford et al. (2013) and references therein), but only that, once appended with this kind of uncertainty, standard assumptions such as RCBR may provide an effective *as if* model of how timing impacts individuals' strategic reasoning.

phenomenon can be found in WY's space.[5]

The rest of the paper is organized as follows: Section 1.1 presents a leading example; Section 2 introduces the model; Section 3 formalizes the notion of RCBR under extensive-form uncertainty; Section 4 contains our main result, Theorem 1. In Section 5 we explore some of Theorem 1's implications for eductive coordination and robust refinements, as well as variations with one-sided uncertainty. Section 6 presents the key steps of the proofs of our main results, as well as some extensions. Section 6.2 may be of independent interest, in that it contains a general result on the structure of rationalizability for static games with arbitrary spaces of payoff uncertainty (with or without richness) and general information partitions. Section 7 reviews the most closely related literature and concludes.

1.1 Leading Example

We begin with a simple example to illustrate the basic elements of our model and some of our results. Consider the following 'augmented' Battle of the Sexes:

	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 have conflicting preferences over which equilibrium they would like to coordinate on. Clearly, if it's common knowledge that the game is static, everything is rationalizable (and, hence, consistent with RCBR).

Now, suppose that players commonly agree that player 1 chooses earlier than 2, but there is uncertainty over whether his action will be observed by 2. We let ω^0 denote the state of the world in which actions are not observable, and let ω^1 denote the case in which 1's action is observable. If the true state is ω^1, and this is common knowledge, the only strategy profile consistent with RCBR is the backward induction solution, which induces 1's favorite equilibrium outcome, (U, L). Imagine next a situation in which the game is actually static (i.e., the true state is ω^0), and both players know it, but 2 thinks that 1 thinks it common belief that the state is ω^1. Then, 2 expects 1 to choose U, and hence choosing L is 2's only best reply. Moreover, if 1 believes that 2's beliefs are just as described, she also picks U as the only action consistent with RCBR. But then, if 2 believes the above, his unique best response is to indeed play L, and so on. Iterating this argument, one can see that 1 and 2 may share arbitrarily many levels of mutual belief that the game is static, and yet have (U, L) as the only outcome consistent with RCBR. Thus, 1 de facto has a first-mover advantage, if

[5] The existence of an u.h.c. refinement of R could perhaps be expected since, as our proof shows, our exercise can be mapped to one of payoff uncertianty, just *without* richness (Penta (2013), however, cautioned against too hasty conclusions of this sort, providing sufficient conditions for WY's unrefinability result without richness). But the fact that both R and RP generically coincide and are single-valued is *not* a natural implication of the lack of richness, which per se often implies open sets of types with multiple rationalizable actions.

she is merely believed to have it at some arbitrarily high order of beliefs. Proposition 3 in Section 5 implies that, if the only uncertainty concerns the observability of 1's action, then this selection occurs for a generic set of belief hierarchies in this game. In this sense, 1's first-mover advantage is *pervasive*, regardless of the actual observability of her action.

Clearly, if we considered symmetric uncertainty, and also included a state ω^2 in which it is 1 who observes 2's action, a similar argument would uniquely select (M, C). Hence, with two-sided uncertainty, no player would necessarily obtain a first-mover advantage, but it can still be shown that no open set of belief hierarchies would select actions D and R. Proposition 2 in Section 5 shows that, for a class of games which includes this example, the predictions consistent with RCBR generically select components of the Stackelberg profiles.

By the same logic, if payoffs were such that the Stackelberg outcomes coincided (which would be the case, for instance, in stag-hunt, in pure coordination games, but also in zero-sum games with pure equilibria), then the Stackelberg profile would be the only outcome consistent with RCBR for a generic set of belief hierarchies, thereby implying equilibrium coordination on the basis of RCBR alone. That is the logic of Proposition 1 in Section 5. (Comparisons between our results and WY's will be discussed in Sections 4 and 6)

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, as let as usual $A := A_1 \times A_2$. Similar to the 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 ω^0 represents the state in which the game is actually static; ω^i represents the state in which the game has perfect information, with player i moving first. (Some extensions are discussed in Section 6.4.) We maintain throughout 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 $A_i' \subseteq A_i$,* $\left|\arg\max_{a_i \in A_i'} u_i^*(a_i, a_j^*(a_i))\right| = 1$.

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 ω^1 and ω^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*, or *Stackelberg, outcome* in the game in which ω^i is common knowledge. We will also refer to a_i^i as i's *Stackelberg action*.

Information: As in Robson (1994), there are two possible pieces of 'hard information' for a player: either he knows he plays knowing the other's action (he is 'second', θ_i''), or not (denoted by θ_i'). We let $\Theta_i = \{\theta_i', \theta_i''\}$ denote the set of *information types*, generated by the information partition over Ω 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 $\omega = \omega^1$. In short, letting $\theta_i(\omega)$ denote the cell of i's information partition which contains ω, we have $\theta_i(\omega) \cap \theta_j(\omega) = \{\omega\}$ for all $\omega \in \Omega$.

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

As usual, any type in a (consistent) type space induces a belief hierarchy over Ω.[6] 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 ω. Finally, we say that type t_i is *finite* if it belongs to a finite belief-closed subset of T^*.

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 Θ_i). With a slight abuse of notation, we can thus write $S_i(t_i)$ to refer to $S_i(\omega)$ such that $\omega \in \hat{\theta}_i(t_i)$, and we let $S_i := \bigcup_{\omega \in \Omega} S_i(\omega)$. For any $\omega \in \Omega$, let $u_i(\cdot, \omega) : S(\omega) \to \mathbb{R}$ be such that:

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

3 Rationality and Common Belief in Rationality

We are interested in the behavioral implications of players' Rationality and Common Belief in Rationality (RCBR) in this setting, where "rationality" is understood in the sense of *sequential rationality* for types t_i with information $\hat{\theta}_i(t_i) = \theta_i''$. Under Assumption 1, these ideas can be expressed in the interim strategic form, by letting types 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. Formally: fix a type space $\mathcal{T} = (T_i, \hat{\theta}_i, \tau_i)_{i=1,2}$; for any i, and t_i, let $R_i^0(t_i) = S_i(t_i)$ and $R_j^0 = \{(s_j, t_j) : s_j \in R_j^0(t_j)\}$. Then, in the first round, types who move second delete all weakly dominated strategies (to capture the idea of *sequential* rationality); all other types instead only delete strictly dominated strategies: For each $t_i \in T_i$,

$$R_i^1(t_i) := \left\{ s_i' \in R_i^0(t_i) : \begin{array}{r} \exists \mu \in \Delta\left(R_j^0 \times \Omega\right) \text{ s.t.: } (i) \text{ marg}_{T_j \times \Omega} \mu_i = \tau_i(t_i) \\ (ii)\ s_i' \in \arg\max_{s_i \in S_i(t_i)} \sum_{\omega \in \hat{\theta}_i(t_i)} \sum_{s_j \in S_j(\omega)} \mu_i[(s_j, \omega)] u_i(s_i, s_j, \omega)) \\ (iii) \text{ if } \theta_i(t_i) = \theta_i'' \text{ and } \mu[(s_j, t_j)] > 0, \\ \text{then } \mu[(s_j', t_j)] > 0, \forall s_j' \in S_j(t_j) \end{array} \right\}.$$

[6]The consistency requirement restricts such hierarchies to be consistent with the type's information, but the construction of the universal type space is standard from Brandenburger and Dekel (1993).

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

$$R_i^k(t_i) := \left\{ s_i' \in R_i^0(t_i) : \begin{array}{l} \exists \mu \in \Delta\left(R_j^{k-1} \times \Omega\right) \text{ s.t.: } (i) \operatorname{marg}_{T_j \times \Omega} \mu_i = \tau_i(t_i) \\ (ii)\ s_i' \in \arg\max_{s_i \in S_i(t_i)} \sum_{\omega \in \hat{\theta}_i(t_i)} \sum_{s_j \in S_j(\omega)} \mu_i[(s_j, \omega)] u_i(s_i, s_j, \omega)) \end{array} \right\},$$

and let $R_i(t_i) := \bigcap_{k \geq 0} R_i^k(t_i)$. 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) can be used to show that $R_i(t_i)$ characterizes the behavioral implications of RCBR, given t_i's beliefs. This solution concept is best thought of as a form of extensive-form rationalizability, with the proviso that types in our type spaces may be uncertain over the extensive form.[7]

Example 1 Consider a type space with types $T_i = \left\{t_i^1, t_i^0, t_i^2\right\}$ for each $i = 1, 2$, where 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 knows that he is not second, and attaches probability p to (t_j^0, ω^0) and $(1-p)$ to (t_j^i, ω^i). Hence, if $p = 1$, t_i^0 represents common belief in the static game; but for $p \in (0, 1)$, t_i^0 is uncertain whether he is part of a static game or the first-mover in a dynamic game. Formally, the type space is such that $\omega^x \in \hat{\theta}_i(t_i^x)$ for each $x = 0, 1, 2$; $\tau_i(t_i^x)[(t_j^x, \omega^x)] = 1$ if $x = 1, 2$, whereas $\tau_i\left(t_i^0\right)[(t_j^0, \omega^0)] = p$ and $\tau_i\left(t_i^0\right)[(t_j^i, \omega^i)] = 1 - p$.

Now consider the example in Section 1.1. Clearly, we have $a^1 = (U, L)$, $a^2 = (M, C)$, and in the following we let $a' = (D, R)$. First note that $S_i\left(t_i^i\right) = S_i\left(t_i^0\right) = A_i$ and $S_j(t_j^i) = A_j^{A_i}$. Since no action is dominated for t_i^i, $R_i^1\left(t_i^i\right) = A_i$, whereas the only non-weakly dominated strategy for t_j^i is its reaction function: $R_j^1(t_j^i) = \{a_j^*(\cdot)\}$. 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 $a^i = (a_i^i, a_j^*\left(a^i\right))$. If $p = 1$, it is also easy to check that $R_i\left(t_i^0\right) = A_i$, as in standard (static) rationalizability (Bernheim (1984) and Pearce (1984)).

If $p \in (0, 1)$, 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 against type t_j^i, which would observe i's action. 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 t_j^i, who plays $a_j^*(\cdot)$, and hence play their Stackelberg action a_i^i: $R_i^2\left(t_i^i\right) = R_i\left(t_i^i\right) = \left\{a_i^i\right\}$, $R_i^1(t_i^j) = R_i(t_i^j) = \{a_i^*(\cdot)\}$. Type t_i^0 thinks that, with probability $(1-p)$, he faces t_i^j (who plays $a_j^*(\cdot)$), otherwise 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 μ over t_j^0's action,

[7] In fact, when the extensive form is common knowledge, it can be shown that in environments with private values, Penta's (2012) interim sequential rationalizability (which is a version of extensive-form rationalizability) is equivalent to applying Dekel and Fudenberg's (1990) procedure to the interim normal form, and to Ben-Porath's (1997) solution concept in games with complete information. In two-stage games with complete and perfect information with no relevant ties, all these concepts yield the backward induction solution.

is therefore to choose $a_i' \in A_i$ that maximizes the following expected payoff:

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

Hence, $R_i^2\left(t_i^0\right) = \{a_i \in A_i : \exists \mu \in \Delta\left(A_j\right) \text{ s.t. } a_i \in \arg\max_{a_i' \in A_i} EU_i\left(a_i'; p, \mu\right)\}$, that is:

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

□

The combination of static and dynamic best-responses illustrated in this example will play a central role in our analysis, since the behavior of the R_i correspondence around the natural benchmarks (i.e., the types which commonly believe ω^0 and ω^i) will in general depend on its solutions for other belief hierarchies, including those in which players are uncertain over whether the game is static or not. We present next two robustness properties of R_i:

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 $(\hat{\theta}_i\left(t_i\right), \hat{\pi}_i(t_i)) = (\hat{\theta}_i(\tilde{t}_i), \hat{\pi}_i(\tilde{t}_i))$, then $R_i(t_i) = R_i(\tilde{t}_i)$.*

Lemma 1 ensures that the predictions of R_i only depend on a type's information and belief hierarchy, not on the particular type space used to represent it. It thus enables us to study R_i as a correspondence on the universal type space, $R_i : T_i^* \rightrightarrows S_i$.[8]

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

This result shows that, similar to ICR and ISR on the universal type space generated by a space of payoff uncertainty, R_i is u.h.c. on our universal type space. This is a robustness property in that it ensures that anything that is ruled out by R_i for some type $t_i \in T_i^*$, is also ruled out for all types in a neighborhood of t_i. This is an important property in the above mentioned literature. For instance, WY's unrefinability results (respectively, Penta's (2012)) can be summarized by saying that ICR (resp., 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., with the extensive-form uncertainty we consider here it will not be the *strongest* u.h.c. solution concept on T_i^*: a proper refinement of R_i is also u.h.c, and hence 'robust' in our space.

4 Robust Predictions: Characterization

In this section we characterize the strongest predictions consistent with RCBR that are robust to higher order uncertainty over the observability of actions. We begin by constructing a set of

[8] This is a standard property for solution concepts with correlated conjectures, such as Dekel et al.'s (2006, 2007) interim correlated rationalizability (ICR) and Penta's (2012) interim sequential rationalizability (ISR).

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 the example in Section 1.1. There, an 'infection argument' showed that the uniqueness of the backward induction solution for types that commonly believe in ω^i propagates to types sharing n levels of mutual belief in ω^0 through a chain of unique best replies. This type of argument is standard in the literature, and it generally involves 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 WY, 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 WY's in our analysis: first, only the backward induction outcomes can serve as seeds; second, best responses must account for the 'hybrid' problems illustrated in Example 1. 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 $\mathcal{B}_i^1 := \{a_i^i\}$ and for $k \geq 1$,

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

Since G^* is finite, there exists $m < \infty$ 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 2 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 $a^2 = (M, C)$, and hence $\mathcal{B}_1^1 = \{U\}$, $\mathcal{B}_2^1 = \{C\}$. 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 allowing as many seeds as possible (richness condition), but accounting for sequential best replies. Penta (2013) and Chen et al. (2014) instead keep standard (static) rationality, but relax the richness assumption. The reasons why the current analysis cannot be cast within any of the existing frameworks are explained at the end of this section.

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 $D \in \mathcal{B}_1^3$ and $\mathcal{B}_i = A_i$ for both i. If instead $x = 0$, then it is easy to check that $\mathcal{B}_1 = \{U, M\}$ and $\mathcal{B}_2 = \{L, C\}$. $\square$

We introduce next a solution concept, $RP_i : T_i^* \rightrightarrows A_i$, obtained by applying the same iterated deletion procedure as R_i, but starting from the set $\mathcal{B}$ instead of A. Since, under the maintained Assumption 1, it can be shown that $a^*(\cdot)$ is the only strategy that is not weakly dominated for all types that move second, it is convenient to initialize the procedure directly from this point. Formally: for each i and t_i, let

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

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

$$RP_i^k(t_i) := \left\{ s_i' \in RP_i^0(t_i) : \begin{array}{l} \exists \mu \in \Delta\left(RP_j^{k-1} \times \Omega\right) \text{ s.t.: } (i) \operatorname{marg}_{T_j \times \Omega} \mu_i = \tau_i(t_i) \\ (ii)\ s_i' \in \arg\max\limits_{s_i \in S_i(t_i)} \sum_{\omega \in \hat{\theta}_i(t_i)} \sum_{s_j \in S_j(\omega)} \mu_i[(s_j, \omega)] u_i(s_i, s_j, \omega)) \end{array} \right\},$$

and $RP_i(t_i) := \bigcap_{k \geq 0} RP_i^k(t_i)$. 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.

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 $(\theta_i(t_i), \hat{\pi}_{i,k}(t_i)) = (\theta_i(t_i'), \hat{\pi}_{i,k}(t_i'))$, then $RP_i^k(t_i) = RP_i^k(t_i')$.*

(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 $(t_i^\nu)_{\nu \in \mathbb{N}}$ in T_i^* 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, $RP_i(\cdot)$ is the *strongest* robust refinement of R_i, since no refinement of $RP_i(\cdot)$ is u.h.c.; (ii) second, R_i and RP_i generically coincide on the universal type space, and deliver the same unique prediction – hence, not only is $RP_i(\cdot)$ a strongest u.h.c. refinement of $R_i(\cdot)$, but it also 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 u.h.c., 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 corollary:

Corollary 1 *(i) No proper refinement of RP_i is upper hemicontinuous on T_i^*.*
(ii) R_i coincides with *RP_i and is* single-valued *over an open and dense subset of T_i^*.*
(iii) For any t_i, if there exists a sequence $(t_i^\nu)_{\nu\in\mathbb{N}}$ in 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 WY, the implications are very different: higher order uncertainty over the observability of actions supports a *robust refinement* of R. 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 3 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

If players commonly believe in ω^0, the rationalizable set for this game is $R(t^{CB}(\omega^0)) = \{M, D\} \times \{C, R\}$. The Stackelberg profiles are $a^1 = (U, L)$ and $a^2 = (M, C)$, and it is easy to check that $\mathcal{B} = \{U, M\} \times \{L, C\}$, and hence $RP(t^{CB}(\omega^0)) = R(t^{CB}(\omega^0)) \cap \mathcal{B} = \{(M, C)\}$. □

The result that R_i and RP_i generically coincide (part (i) of Corollary 1) 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, and hence discard the actions which are only rationalizable for non-generic sets of types. Then, part (i) of Corollary 1 implies that whereas RCBR may deliver less sharp predictions than $RP(\cdot)$ for non-generic types (such as $t^{CB}\left(\omega^0\right)$ 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 'regular predictions' of RCBR. Formally:

Definition 1 *Action a_i is a* regular prediction of RCBR *for type t_i if $a_i \in R_i(t_i)$ and for any neighborhood $\mathcal{N}(t_i)$ of t_i, there exists an open set $U \subset \mathcal{N}(t_i)$: $a_i \in R_i(t_i')$ for all $t_i' \in U$.*

Corollary 2 *Action a_i is a* regular prediction of RCBR *for t_i if and only if $a_i \in RP_i(t_i)$.*[10]

The key steps of the proof of Theorem 1, which we discuss in Section 6, consist of re-casting the problem of uncertainty over the information structure, as one of payoff uncertainty of an *auxiliary static game* (Section 6.1). This way, we can obtain the main result using techniques which are closer to the literature on payoff uncertainty, and hence easier to compare and possibly extend to other directions (see Section 6.4).[11] The auxiliary game, however, does not satisfy WY' nor Penta's (2012) richness conditions, and it must account for players' information partition over the space of uncertainty. Thus, none of the existing results can be directly applied to the auxiliary game. Our Lemma 4 overcomes this difficulty, by generalizing the analysis in Penta (2013) – which studied the ICR correspondence with payoff uncertainty without richness and without information types – to static games with arbitrary payoff uncertainty (with or without richness) and general information partitions. Hence, while we developed it as a step towards the proof of Theorems 1, this Lemma has intrisic interests in that it generalizes existing results on static games with payoff uncertainty.

5 Applications

In Example 3, not only are the robust predictions particularly sharp, but they also imply that, for a generic set of types with information θ_i', equilibrium coordination arises as the only behavior consistent RCBR, i.e. without imposing correctness of beliefs. In Section 5.1 we consider classes of games in which the robust predictions take this especially strong form, and so for a generic set of types equilibrium coordination arises purely from individual reasoning. Section 5.2 explores other classes of games, in which Theorem 1 also has strong implications, which may or may not lead to eductive coordination. In Section 5.3 we present results for environments with one-sided uncertainty.

5.1 '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, the dominant approach has developed Schelling's (1960) idea of *focal points* (e.g., Sugden (1995), Iriberri and Crawford (2007), etc.), which shifts the discussion on the mechanisms that bring about coordination to

[10] We note that the open sets U in Definition 1 are not required to include t_i. If they did, regularity would be equivalent to lower hemicontinuity, which neither R_i nor RP_i satisfy.

[11] An earlier draft of this work contained a more involved proof which dealt directly with extensive form uncertainty. We thank an anonymous referee for the suggestion to pursue this proof strategy.

external, non mathematical properties of the game.[12] The next result shows that there is an interesting class of games for which higher order uncertainty over the extensive form provides a purely eductive mechanism for equilibrium coordination, based on classical game theoretic assumptions (namely, RCBR), without appealing to any external theory of focal points:[13]

Proposition 1 (Generic Coordination) *For any G^* which satisfies Assumption 1 and in which the two Stackelberg actions coincide ($a^1 = a^2 \equiv \bar{a}$), there exists an open and dense subset $T' \subseteq T^*$ such that, for all $t \in T'$, $\bar{a}$ is the only outcome induced by $R(t)$.*

Note that, since by definition a_j^i is a best response to a_i^i, the condition $a^1 = a^2 \equiv \bar{a}$ implies that $\bar{a}$ is a Nash equilibrium. Hence, Proposition 1 implies 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. While the result follows immediately from Theorem 1, and from the observation that $\mathcal{B} = \{\bar{a}\}$ if $a^1 = a^2 \equiv \bar{a}$, the interest of this proposition is due to the fact that important and seemingly disparate classes of games (which include, for instance, archetypal models of both common interest and pure conflict situations) satisfy the condition $a^1 = a^2$:

Remark 1 *If G^* satisfies Assumption 1, then the condition $a^1 = a^2 \equiv \bar{a}$ holds if G^* belongs to any of the following classes of games:*

1. Coordination games with a unique Pareto efficient equilibrium, $\bar{a}$.

2. Common interest games (cf. Aumann and Sorin (1989)).[14]

3. Zero-sum games with a pure Nash equilibrium, $\bar{a}$.

Proposition 1 is also interesting from the viewpoint of equilibrium refinements. For instance, in *common interest games*, 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, and in any case always relying on the observability of the opponent's actions.[15] In contrast, our efficient coordination result holds for a generic subset of the universal type space, regardless of whether players' actions are actually observable, and as the only outcome consistent with RCBR for those types.

For *zero-sum games*, this result bridges a gap between RCBR and the maxmin solution which has long been discussed in the literature. To illustrate the point, we adapt arguments

[12] 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 [...]."

[13] Alaoui and Penta (2019) provide an alternative mechanism for eductive coordination, based on the endogenous level-k model of Alaoui and Penta (2016).

[14] Formally, 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 (that is, 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). A *common interest* game is a coordination game which also satisfies $u_1^*(a) = u_2^*(a)$ for all $a \in A$

[15] Aumann and Sorin (1989), for instance, support the efficient equilibrium in this very 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.

from Luce and Raiffa (1957) to the following example:[16]

Example 4 Consider the following game, in which $\varepsilon > 0$:

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

First note that: (i) everything is rationalizable in this game; (ii) (M, C) is the maxmin solution; and (iii) $\mathcal{B} = \{(M, C)\}$. 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 [C]. Certaintly (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 [ε]." (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 ε. 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, reflected in the fact that all actions are rationalizable in the example, clashes with the sharpness of the maxmin criterion. On the other hand, the latter is grounded on a simple, if extreme, decision theoretic principle. A classical argument to reconcile the two views is to note that the maxmin action ensures expected utility maximization in the eventuality that one's action is leaked to the opponent (see, e.g. Von Neumann and Morgenstern, 1947).[17] The logic behind our result is reminiscent of that argument. We point out, however, that whereas the standard 'fear of leaks' argument can be thought of as a first-order beliefs effect, Proposition 1 implies that the maxmin action is the only *regular prediction* of RCBR everywhere on T^*, including for types which share arbitrarily many (but finite) orders of mutual belief that leaks have *zero* probability.

The role of the $a^1 = a^2 \equiv \bar{a}$ condition in Proposition 1 is to ensure that $\mathcal{B} = \{\bar{a}\}$, which in turn implies that RP_i is single-valued also at the static common-belief type $t_i^{CB}(\omega^0)$, yielding the eductive coordination result. As shown by Example 3, however, eductive coordination is

[16]Luce and Raiffa's original argument refers to a game that violates Assumption 1, but it applies unchanged to our example, which satisfies Assumption 1.

[17]Robson (1994) and Reny and Robson (2004) have formalized these arguments explicitly, but within a common prior model with uncertainty over the monitoring structure and using equilibrium as solution concept. Nonetheless, many of our insights – including the reconciliation between expected utility and maxmin criterion in zero-sum games – are reminiscent of theirs. But the logic of the results are very different. As we discuss shortly, besides being based on a non-equilibrium concept, our results do not require any first-order uncertainty over the observability of actions. In contrast, and consistent with the classical view, the common prior models in Robson (1994) and Reny and Robson (2004) also introduce uncertainty at the first order of beliefs.

possible even if $a^1 \neq a^2$: all is needed is for RP to uniquely select a Nash equilibrium, which can be ensured for instance if the game is such that, as in Example 3, $\mathcal{B} \cap R_i\left(t_i^{CB}\left(\omega^0\right)\right) = \{\bar{a}\}$ for some Nash equilibrium $\bar{a}$. Various restrictions on payoffs could yield this property. We focused on the $a^1 = a^2$ condition because of its special significance, as discussed.

5.2 Stackelberg Selections

The next result follows from Theorem 1, for a class of games which includes the example in Section 1.1, as well as Harsanyi (1981) and Kalai and Samet's (1984) unanimity games:[18]

Proposition 2 *If G^* satisfies Assumption 1, both players are indifferent over non-Nash equilibrium profiles, and they strictly prefer any Nash-equilibrium to any non-Nash equilibrium profile, then there is an open and dense set $T_i' \subseteq T_i^*$ such that, for any $t_i \in T_i'$, $R_i(t_i) = RP_i(t_i) \in \{\{a_i^i\}, \{a_i^j\}\}$ if $\hat{\theta}_i(t_i) = \theta_i'$ and $R_i(t_i) = \{a^*(\cdot)\}$ if $\hat{\theta}_i(t_i) = \theta_i''$.*

The result follows from the observation that, in games which satisfy the conditions in the proposition, a^i and a^j are Nash equilibria and $\mathcal{B}_i = \{a_i^i, a_i^j\}$. This, together with the fact that $RP_i = R_i$ generically on T^* (Corollary 1), implies the result. Note that the statement of Proposition 2 does not only refer to the neighborhood of the benchmark static types $t_i^{CB}(\omega^0)$, but to the generic predictions of RCBR. Thus, for instance, whereas inefficient equilibrium actions are consistent with RCBR when ω^0 is common belief, generically they are not:

Corollary 3 *In any game which satisfies the conditions in Proposition 2, actions associated to inefficient Nash equilibria are generically ruled out by RCBR.*

5.3 One-sided Uncertainty and Pervasiness of First-Mover Advantage

In this section we consider the implications of maintaining common knowledge that one of the two player's actions is *not* observable, so that the higher order uncertainty only refers to the observability of one of the players' actions. Such one-sided uncertainty is relevant, for instance, if players' choices are irreversible and made with a commonly known order, so that the earlier mover cannot observe the later mover's action; or if players commonly agree that only one of them has successfully committed to ignoring the other player's choice, or that only the actions of one player are effectively irreversible; etc.

Formally, let player 1 denote the player who is commonly known to *not* observe the opponent's action, and consider the smaller space of uncertainty $\Omega^\dagger = \left\{\omega^0, \omega^1\right\}$ (only player 2 knows that state), and let $T_i^\dagger$ denote the universal type space generated by $\Omega^\dagger$. For each i, define the subset of actions $\mathcal{B}_i^\dagger := \bigcup_{k\geq 1} \mathcal{B}_i^k$, where $\mathcal{B}_1^{\dagger,1} := \{a_1^1\}$, $\mathcal{B}_2^{\dagger,1} := \emptyset$ and for each $k \geq 1$:

$$\mathcal{B}_1^{\dagger,k+1} := \mathcal{B}_1^{\dagger,k} \cup \left\{ \begin{array}{l} a_1 \in A_1 : \exists \mu \in \Delta(\mathcal{B}_2^{\dagger,k}), \exists p \in [0,1] \text{ s.t.} \\ \{a_1\} = \arg\max_{a_1' \in A_1} \left(p \cdot \sum_{a_2 \in A_2} \mu[a_2] \cdot u_1^*(a_1', a_2) + (1-p) \cdot u_1^*(a_1', a_2^*(a_2'))\right) \end{array} \right\}$$

[18] Formally, a *unanimity* game is a coordination game (cf. footnote 14) such that, $\forall i$, $u_i^*(a') = u_i^*(a'')$ for all non-Nash equilibrium profiles a', a''.

$$\mathcal{B}_2^{\dagger,k+1} := \mathcal{B}_2^{\dagger,k} \cup \left\{ a_2 \in A_2 : \exists \mu \in \Delta(\mathcal{B}_1^{\dagger,k}) \colon \ \{a_2\} = \arg\max_{a_2' \in A_2} \sum\nolimits_{a_1 \in A_1} \mu[a_1] \cdot u_2^*(a_2', a_1) \right\}$$

Note that $\mathcal{B}_i^\dagger$ is basically the same as the set $\mathcal{B}_i$ defined in Section 2, except that only a_1^1 is taken as a 'seed', not a_2^2. For each i, we define correspondence $RP_i^\dagger$, which is obtained replacing the sets $\mathcal{B}_i$ with $\mathcal{B}_i^\dagger$ in the definition of $RP_i(t_i)$, for each $t_i \in T_i^\dagger$. The next result, analogous to Theorem 1, implies that on this space of uncertainty $RP_i^\dagger$ is both the strongest u.h.c. refinement of R_i and it characterizes its *regular* predictions:

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

The following corollary states properties of $RP_i^\dagger$ analogous to those of Corollaries 1-2:

Corollary 4 *For any player i, the following holds:*
(i) No proper refinement of $RP_i^\dagger$ is upper hemicontinuous on $T_i^\dagger$.
(ii) R_i coincides with $RP_i^\dagger$ and is single-valued over an open and dense set of types $T_i' \subseteq T_i^\dagger$.
(iii) For any $t_i \in T_i^\dagger$, if there exists a sequence $(t_i^\nu)_{\nu\in\mathbb{N}}$ in $T_i^\dagger$ with limit t_i s.t. $R_i(t_i^\nu) = \{s_i\}$ for all $\nu \in \mathbb{N}$, then $s_i \in RP_i^\dagger(t_i)$.
(iv) For any $t_i \in T_i^\dagger$, a_i is a regular prediction of RCBR *for type t_i if and only if $a_i \in RP_i^\dagger(t_i)$.*

This result has especially strong implications in games in which a^1 is also a Nash equilibrium, which is a larger class of games than those considered in Propositions 1 and 2:

Proposition 3 (Pervasiveness of First-Mover Advantage) *If G^* satisfies Assumption 1 and a^1 is one of its Nash-equilibria, then there is an open and dense subset of types $T_i' \subseteq T_i^\dagger$ such that, for all $t_i \in T_i'$, $R_i(t_i) = \{a_i^1\}$ if $\hat{\theta}_i(t_i) = \theta_i'$, and $R_i(t_i) = \{a^*(\cdot)\}$ if $\hat{\theta}_i(t_i) = \theta_i''$.*

Hence, in this class of games, the presence of a state in which 1 has a first-mover advantage, implies that 1 has a *de facto* first-mover advantage generically on $T_i^\dagger$. In this sense, we say that a first-mover advantage is *pervasive*, and it arises (generically) independent of the actual observability of 1's actions, also for types who share arbitrarily many (but finite) orders of mutual beliefs that 1's action is *not* observable.[19]

The result has important strategic implications, in that it points at the impact that mechanisms to establish common knowledge of one-sided uncertainty may have in the presence of this kind of higher order uncertainty. Various kinds of mechanisms may produce such one-sided uncertainty: for instance, a commonly known order of (irreversible) moves; environments with simultaneous choices with (at most) one-sided reversibility; one-sided ability

[19] The message of Proposition 3 may appear to be in sharp contrast with Bagwell (1995), who argued that the first-mover advantage is rather fragile. Aside from the use of a common prior model, the most important difference is that the information at states $(\omega^i)_{i=1,2}$ violates Bagwell's identical support assumptions on the distributions of signals under different actions. Also, Bagwell (1995) considers games which do not fall within the scope of Proposition 3. For such games, the first-mover advantage may not be 'pervasive', but it would still be uniquely selected in an open neighborhood of $t^{CB}(\omega^1)$, and hence locally robust in our model.

to commit to not observing others' actions, etc. All these mechanisms are formally equivalent, since they all entail a space with one-sided uncertainty $\Omega^{\dagger}$. The commonly known *timing* of moves, however, is perhaps the simplest and most obvious to consider. Within this context, the result suggests that – by determining the direction of the one-sided uncertainty – *timing* of moves (plus irreversibility) may determine the attribution of the strategic advantage, independent of the actual observability of actions.

The notion that timing and commitment has strategic importance, beyond actual observability, has been discussed by Kreps (1990), and the idea has received strong support by the experimental literature, which has shown for instance that asynchronous play in the Battle of the Sexes drastically affects subjects' behavior, in that it induces coordination on the earlier mover's Stackelberg profile, even when his action is *not* observable (see, e.g., Camerer (2003) and references therein). 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 3 is the first result to make sense of this solid experimental evidence, without appealing to behavioral theories or notions of bounded rationality, while maintaining non-observability of actions and without extending the game under consideration.[20]

Finally, note that the result in Proposition 3 implies that, with one-sided uncertainty, higher order uncertainty over the observability of actions yields eductive coordination even in games which do not satisfy the condition of Proposition 1.

6 Proof of the Main Results

In this section we explain the key ideas in the proofs of the main results (particularly of parts (*iii*) in Theorems 1 and 2), as we skecthed at the end of Section 4: the auxiliary game is introduced in Section 6.1; the main lemma on general payoff uncertainty is Section 6.2; its application to Theorems 1 and 2 is in Section 6.3. Section 6.4 discusses possible extensions.

6.1 The Auxiliary Game

Starting from our baseline game $G^* = (A_i, u_i^*)_{i=1,2}$, we define the *auxiliary game* as a static game with payoff uncertainty $\hat{G} = (\Omega, (\hat{A}_i, \hat{u}_i)_{i=1,2})$, where $\hat{A}_i = A_i \cup \{\hat{a}_i\}$ and, letting $M \in \mathbb{R}$ is such that $M > \max_{i \in I} \max_{a \in A} |u_i^*(a)|$, $\hat{u}_i : \hat{A}_1 \times \hat{A}_2 \times \Omega \to \mathbb{R}$ is such that:

$$\hat{u}_i(a,\omega) = \begin{cases} u_i^*(a) & \text{if } a \in A, \\ -M & \text{if } a_i = \hat{a}_i \text{ and } \omega \neq \omega^j, \\ 2M & \text{if } a_i = \hat{a}_i \text{ and } \omega = \omega^j, \\ u_i^*(a_i, a_j^*(a_i)) & \text{if } a_j = \hat{a}_j,\ a_i \in A_i \text{ and } \omega = \omega^i, \\ M & \text{if } a_j = \hat{a}_j,\ a_i \in A_i \text{ and } \omega \neq \omega^i. \end{cases}$$

Note that, by construction, the 'extra' action $\hat{a}_i$ is strictly dominant at ω^j and dominated at states $\omega \neq \omega^j$. The rest of the actions yield the same payoffs as in G^* at states $\omega \neq \omega^i$ when the

[20]Amershi, Sadanand and Sadanand (1992) developed solution concepts that assign a specific role to timing as a coordinating device, and hence they appeal to 'external' considerations. Hammond (2008) obtains the Stackelberg outcome as a refinement of the static game appended with a one-sided cheap talk stage.

opponent chooses a standard action $a_j \in A_j$, and M if he chooses $\hat{a}_j$. Information partitions over Ω are the same as in Section 2, with associated universal type space T^* ($T^\dagger$ denotes the universal type space associated to the space with one-sided uncertainty $\Omega^\dagger$ of Section 5.3). For any $t_i \in T_i^*$ (resp., $T_i^\dagger$), we let $\mathrm{ICR}_i(t_i)$ and $\mathrm{ICR}_i^k(t_i)$ denote, respectively, the set of interim correlated rationalizable actions (Dekel et al., 2007) and the set of actions which survived the k-th round of deletion in the auxiliary game $\hat{G}$ for type t_i. $R_i(t_i)$ denotes the solution concept defined in Section 3 for the underlying game with extensive form uncertainty.

Lemma 3 *For any $t_i \in T_i^*$ (resp., $T_i^\dagger$) , and for any $k \geq 1$: if $\hat{\theta}_i(t_i) = \theta_i'$, then $R_i^k(t_i) = \mathrm{ICR}_i^k(t_i)$; if $\hat{\theta}_i(t_i) = \theta_i''$, then $R_i^k(t_i) = \{a_i^*(\cdot)\}$ and $\mathrm{ICR}_i^k(t_i) = \{\hat{a}_i\}$.*

This lemma clarifies that the R_i correspondence for the game with extensive-form uncertainty, can be equivalently characterized by the ICR_i correspondence for the auxiliary game, provided that one identifies the dummy action $\hat{a}_i$ with what would be i's sequential best response in state ω^j. This lemma also implies Lemmas 1 and 2.

6.2 General Payoff Uncertainty with Information Partitions

Let $\hat{G} = (\Omega, (\Theta_i, \hat{A}_i, \hat{u}_i)_{i=1,2})$ be an arbitrary finite two-person normal-form game with incomplete information, where Ω is an arbitrary set of payoff states, and $(\Theta_i)_{i=1,2}$ is a profile of information partitions over Ω. For any information-based type space $(T_i, \hat{\theta}_i, \tau_i)_{i=1,2}$ where $\tau_i : T_i \to \Delta(T_j \times \Omega)$ is s.t. $\tau_i(t_i)[\{(t_j, \omega) \in T_j \times \Omega : \omega \in \hat{\theta}_i(t_i) \cap \hat{\theta}_j(t_j)\}] = 1$, and for any t_i, we let $\mathrm{ICR}_i(t_i)$ the set ICR actions for type t_i. We define the set $\mathcal{C}_i(\theta_i)$ of actions which can be 'infected' for an information type, through a chain of unique best responses which traces back to states ω at which some action is strictly dominant, and are also consistent with each type's information. Formally, let $\mathcal{C}_i^0(\theta_i)$ denote the set of actions in $\hat{A}_i$ which are uniquely rationalizable for some t_i with information θ_i (e.g., it includes any action which is strictly dominant for some $\omega \in \theta_i$), and then, define recursively, for every $k \geq 1$,

$$\mathcal{C}_i^k(\theta_i) := \mathcal{C}_i^{k-1}(\theta_i) \cup \left\{ a_i \in \hat{A}_i : \begin{array}{ll} \multicolumn{2}{c}{\exists \mu_i \in \Delta(\hat{A}_j \times \Omega) :} \\ (i) & \mu_i[(a_j, \omega)] > 0 \Rightarrow \omega \in \theta_i \text{ and } a_j \in \mathcal{C}_j^{k-1}(\theta_j(\omega)), \\ (ii) & \arg\max_{a_i' \in \hat{A}_i} \sum_{a_j \in \hat{A}_j} \sum_{\omega \in \Omega} \mu_i[a_j] \cdot \hat{u}_i(a_i', a_j, \omega) = \{a_i\} \end{array} \right\}.$$

Finally, set $\mathcal{C}_i(\theta_i) := \bigcup_{k \geq 0} \mathcal{C}_i^k(\theta_i)$. (The finiteness of $\hat{A}_i$ ensures that the union stabilizes after finitely many iterations.) We next define a refinement of ICR for which we will we show it is possible to perform the nearby selection result a la WY: For each type t_i let $\mathrm{ICR}_i^{\mathcal{C}}(t_i) := \bigcap_{k \geq 0} \mathrm{ICR}_i^{\mathcal{C},k}(t_i)$, where $\mathrm{ICR}_i^{\mathcal{C},0}(t_i) := \mathcal{C}_i(\hat{\theta}_i(t_i))$ and then, for every $k \geq 0$,

$$\mathrm{ICR}_i^{\mathcal{C},k+1}(t_i) := \left\{ a_i \in \hat{A}_i : \begin{array}{ll} \multicolumn{2}{c}{\exists \mu_i \in \Delta(\hat{A}_j \times T_j \times \Omega) : (i) \ \mathrm{marg}_{T_j \times \Omega} \mu_i = \tau_i(t_i)} \\ (ii) & \mu_i \left[\left\{ (a_j, t_j) \in \hat{A}_j \times T_j : a_j \in \mathrm{ICR}_j^{\mathcal{C},k}(t_j) \right\} \times \Omega \right] = 1 \\ (iii) & a_i \in \arg\max_{a_i' \in \hat{A}_i} \sum_{a_j \in \hat{A}_j} \sum_{\omega \in \Omega} \mu_i[(a_j, t_j)] \cdot \hat{u}_i(a_i, a_j, \omega) \end{array} \right\}.$$

Lemma 4 *For any finite type t_i and any strategy $a_i \in \mathrm{ICR}_i^{\mathcal{C}}(t_i)$ there exists a sequence of finite types $(t_i^\nu)_{\nu\in\mathbb{N}}$ with limit t_i and such that $\mathrm{ICR}_i^{\mathcal{C}}(t_i^\nu) = \mathrm{ICR}_i(t_i^\nu) = \{a_i\}$ for any $\nu \in \mathbb{N}$.*

This lemma generalizes the main result in Penta (2013) to settings with general information structures. (Penta's (2013) result obtains for the special case in which each Θ_i is the trivial partition.) In the next subsections, we will apply it to the specific auxiliary game introduced above, for which it turns out that not only ICR coincides with R (Lemma 3), but also that $\mathrm{ICR}_i^{\mathcal{C}}$ coincides, respectively, with the RP and $RP^\dagger$ correspondences when the set of states is Ω and $\Omega^\dagger$, respectively. Part (*iii*) of Theorems 1 and 2 then follow.

6.3 Proof of Part (iii) in Theorems 1 and 2

Applying the definitions of $\mathcal{C}_i^k$ and $\mathrm{ICR}_i^{\mathcal{C},k}$ to the auxiliary game, Lemma 10 in the appendix shows that, for each i, $\mathcal{C}_i(\theta_i^{'}) = \mathcal{B}_i$ and $\mathcal{C}_i(\theta_i^{''}) = \{\hat{a}_i\}$. Similarly, for the $\Omega^\dagger$ space, we obtain $\mathcal{C}_i(\theta_i^{'}) = \mathcal{B}_i^\dagger$ and $\mathcal{C}_2(\theta_2^{''}) = \{\hat{a}_2\}$. Given this, we obtain the following results:

Lemma 5 *For every type $t_i \in T_i^*$ and $k \geq 0$, if $\hat{\theta}_i(t_i) = \theta_i'$ then $RP_i^k(t_i) = \mathrm{ICR}_i^{\mathcal{C},k}(t_i)$ and if $\hat{\theta}_i(t_i) = \theta_i''$, then $RP_i^k(t_i) = \{a_i^*\}$ and $\mathrm{ICR}_i^{\mathcal{C},k}(t_i) = \{\hat{a}_i\}$.*

Lemma 6 *For every type $t_i \in T_i^\dagger$ and $k \geq 0$, if $\hat{\theta}_i(t_i) = \theta_i'$ then $RP_i^\dagger(t_i) = \mathrm{ICR}_i^{\mathcal{C},k}(t_i)$ and if $\hat{\theta}_i(t_i) = \theta_i''$, then $RP_i^\dagger(t_i) = \{a_i^*\}$ and $\mathrm{ICR}_i^{\mathcal{C},k}(t_i) = \{\hat{a}_i\}$.*

Part (*iii*) of Theorem 1 (resp., Theorem 2) follows directly from the combination of Lemmas 3, 4 and 5 (resp., 6).

6.4 Payoff Uncertainty and Other Extensions: Discussion

Thanks to the generality of Lemma 4, our proof strategy can also be adapted to also include payoff uncertainty. In practice, this would boil down to a larger set of states Ω, and information partitions to represent the joint information players may have over payoffs and extensive form. Depending on the situation, this would require alternative and more complex versions of the *auxiliary game* $\hat{G}$, which would crucially depend not only on the extra payoff states, but also on the details of the information structures, which may affect the backward induction solution at different payoff states.[21]

The logic of the lemma, however, remains unchanged, and hence it clarifies that all such questions can easily be addressed via a simple plug-in exercise, through the 'seeds' with which the definition of the $\mathcal{C}_i(\theta_i)$ sets are initialized. As long as the added payoff states satisfy a slight strengthening of Assumption 1, the result would still go through, with the only difference that the sets $\mathcal{C}$ may grow larger (though not necessarily), and hence entail

[21]For instance, if one introduced extra payoff states and modelled payoff structures such that players only know whether they move second or not, but with no information on payoffs, then the auxiliary game would be such that for every payoff state $\hat{\omega}$ which is added, there would be a corresponding ficititious action $\hat{a}_i^{\hat{\omega}}$ which in the auxiliary game plays the role of the sequential best response at payoff state $\hat{\omega}$; at the opposite extreme, if the information partitions were such that players payoff uncertainty only affects the state of the world in which the game is static, then no substantial variation would be needed to the auxiliary game in Section 6.1, beyond the obvious specification of the effects of payoff states on the payoffs of the static game; etc.

weaker strongest robust predictions. For example, if one added a richness condition a la WY, then trivially $\mathcal{C}_i(\theta_i) = \hat{A}_i$, and hence the strongest robust predictions around types $t_i^{CB}(\omega^0)$ would be the same as in WY. Richness, however, often entails an unnecessarily demanding robustness requirement, and the plausibility of considering payoff states which induce new 'seeds' (and, hence, might affect the robust predictions) necessarily depends on the specific application. For instance, suppose that the matrix of the game in Section 1.1 does not represent players' payoffs, but monetary payments, according to some commonly known 'rules of the game' $g : A \to \mathbb{R} \times \mathbb{R}$. The actual payoffs would thus depend on players' Bernoulli utility functions $v_i : \mathbb{R} \to \mathbb{R}$, with $u_i(a) = v_i(g_i(a))$. In such a setting, it certainly make sense to consider uncertainty over utility functions v_i. But in most economic applications it would still be meaningful to maintain common knowledge that such v_i are increasing. Even if we took the space of payoff uncertainty to include all possible profiles of such functions, the sets $\mathcal{C}$ (and, hence, the robust predictions) would still not be affected, because the backward induction outcomes in such a game are uniquely pinned down by ordinal preferences, and no action in that game can be made strictly dominant without violating monotonicity of the v_i, or also relaxing common knowledge of the outcome function g.[22]

The discussion above also applies to extensions of the model with richer possibilities of extensive-form uncertainty. For instance, besides having states in which players observe others' actions perfectly or not at all, one may consider states in which the second mover has partial information about the earlier mover's action. This situation too would boil down to a larger set of states Ω and, depending on the payoff structure and the nature of the monitoring technology at the added states, may require different specifications of the auxiliary game. But once again, the logic of Lemma 4 still applies, and as long as the added states satisfy a strengthening of Assumption 1, the main result goes through unchanged, with the only difference that the sets $\mathcal{C}$ may grow larger (though not necessarily).

In short, our proof can easily be adapted to also accommodate richer specifications of extensive-form as well as payoff uncertainty. What kind of payoff uncertainty is sensible to consider, if at all, can only be evaluated on a case by case basis. Given the nature of the perturbations which are considered relevant, whether they will impact the $\mathcal{C}$ sets and hence reduce the bite of the robust predictions, and whether this would also deliver a generic uniqueness result, will depend on the nature of the perturbations, on the informational assumptions, and on the overall structure of the game's payoffs. In this sense, our analysis may serve as a template to address different questions of extensive form uncertainty, with or without payoff uncertainty of various degrees.

7 Related Literature and Concluding Remarks

In this paper we studied the implications of perturbing common knowledge assumptions on the observability of actions. The closest papers to our work are those which study perturbations of common knowledge assumptions on payoffs, following the seminal paper by

[22]We note that this observation applies to any game which satisfies the conditions of Propositions 1, 2, or 3. Relaxing common knowledge of the rules of the game may be sensible in some settings, but less in others.

Weinstein and Yildiz (2007, WY). WY characterize the correspondence of ICR on the universal type space generated by a space of payoff uncertainty which satisfies a richness condition for static games. The analysis has been extended to dynamic games, which require different richness conditions, by Weinstein and Yildiz (2011, 2016), Chen (2012) and Penta (2012). The latter paper also allows for information information partitions with a product structure. Penta (2013) relaxes the richness condition in static games, and studies sufficient conditions for WY'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. Our key Lemma 4, in particular, generalizes the main result in Penta (2013) to static games with general payoff uncertainty and general information partitions.

A few papers have studied models in which players are uncertain over the observability of actions. Robson (1994), in particular, introduced a refinement for two-player nonzero-sum games ("informationally robust equilibrium") using exactly the same space of uncertainty and information partition as in our model. On a similar vein, Reny and Robson (2004) model a situation in which player's types may be uncertain of whether their action will be observed by the opponent, and study the behavior of equilibria in these settings as the distribution approaches the static benchmark. Both these papers adopt an equilibrium approach in a standard common prior setting. Kalai (2004) introduced a notion of 'extensive robust equilibrium' to denote a profile of choices which remains an equilibrium in a large set of extensive forms, and then shows that as the game becomes large, all equilibria become approximately extensively robust. Like the previous papers, Kalai assumes that there is no higher-order uncertainty over observability among players; only the analyst faces such uncertainty. Zuazo-Garin (2017) introduces incomplete information about the information sets over a game-tree 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 discussed in the previous paragraph.

Our main results show that higher order uncertainty over the observability of actions supports a robust refinement of rationalizability, with several implications in important classes of games, such as: (i) eductive coordination in games in which inverting the order of moves does not affect the Stackelberg profiles; (ii) maxmin selection in zero-sum games with pure equilibria; (iii) Stackelberg selections in a class of coordination games. In environments in which only player 1's actions may be observable, but not player 2's (for instance, because 1 is commonly known to move earlier, or to be the only one whose choices are irreversible, etc.), we showed that, in a class of games which generalizes all the above, RCBR generically selects the equilibrium of the static game which is most favorable to player 1. When such one-sided uncertainty stems from a commonly known order of moves, this result also provides a rational basis for the *Kreps hypothesis* (Kreps, 1990), which maintains that timing and commitment may have strategic importance beyond actual observability of actions – 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. Here it emerges as the *only* behavior consistent with RCBR for a generic set of types.

The breadth of these results suggests that further exploring the problem of extensive-form uncertainty may prove to be a fertile direction for future research. In Section 6.4 we discussed how our proof strategy can be adapted to extend the analysis to environments with payoff uncertainty, and to richer extensive form uncertainties, in which actions may be observed with various degrees of precision. The logic of Lemma 4 suggests that it may also be fruitfully applied to games with more than two players. The most obvious challenge in doing this is represented by the richness of the extensive forms that a larger set of players would entail, and the difficulty of devising an auxiliary game which captures such complexity. From a more applied perspective, it would be interesting to further explore the implications of Theorems 1 and 2 to classes of games not covered by Propositions 1-3 above.

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, while in this paper we pursued a 'local' notion of robustness (similar to WY for payoff uncertainty, and Oury and Tercieux (2012) in mechanism design), other recent work (e.g., Peters (2015), Doval and Ely (2016), Makris and Renou (2018)) have sought to characterize the range of possible equilibrium behaviors which could be generated across a large class of extensive forms which are consistent with some minimal information about the game, thereby pursuing a more 'global' approach to extensive-form robustness (in this sense, closer to the works of Bergemann and Morris (2013, 2016) in games with payoff uncertanty, and Bergemann and Morris (2005, 2009) and Penta (2015) in mechanism design).[23] Similarly, intermediate notions of robustness with payoff uncertainty, which have been put forward in the mechanism design literature (e.g., Ollar and Penta (2017, 2019)), 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 a systematic understanding of this question, but the modeling possibilities are very rich, and suggest many promising directions for future research.

Appendix

A Proof of Theorems 1 and 2

A.1 Proof of Part (i)

We begin with Theorem 1, and proceed by induction on k. The claim holds trivially $k = 0$. Next, suppose it also holds for $k \geq 0$, and fix player i and types $t_i \in T_i$ and $t'_i \in T'_i$ s.t. $(\hat{\theta}_i(t_i), \hat{\pi}_{i,k+1}(t_i)) = (\hat{\theta}_i(t'_i), \hat{\pi}_{i,k+1}(t'_i))$. It suffices to show only one inclusion. Pick $s_i \in RP_i^{k+1}(t_i)$ and conjecture μ_i which justifies the inclusion of s_i in $RP_i^{k+1}(t_i)$. Define measure $\nu_i \in \Delta(S_j \times T_j \times \Omega)$ as follows: $\nu_i[E] = \tau_i(t_i)[\text{Proj}_{T_j \times \Omega}(E)]$ for any measurable $E \subseteq S_j \times T_j \times \Omega$.

[23] Similar to the paper by Kalai (2004) discussed above, in Peters (2015), Doval and Ely (2016) and Makris and Renou (2018) the extensive form is common knowledge among the players, only unknown to the analyst.

Clearly, μ_i is absolutely continuous w.r.t. ν_i, and hence we can pick the corresponding Radon-Nykodym derivative, denoted by f_i. Define $\mu'_i \in \Delta(S_j \times T'_j \times \Omega)$ s.t. $\mu'_i[E] := \int_E f_i \mathrm{d}\nu'_i$ for any measurable $E \subseteq S_j \times T'_j \times \Omega$, where $\nu'_i[E] := \tau_i(t'_i)[\mathrm{Proj}_{T'_j \times \Omega}(E)]$. Clearly, μ'_i justifies the inclusion of s_i in $RP_i^{k+1}(t'_i)$. For Theorem 2, take types t_i and t'_i that induce types in $T_i^\dagger$, and repeat the argument substituting RP for $RP^\dagger$ and Ω for $\Omega^\dagger$.

A.2 Proof of Part (ii)

We complete the proof for Theorem 1 (as above, for Theorem 2 simply substitute RP for $RP^\dagger$ and Ω for $\Omega^\dagger$). Upper-hemicontinuity and nonemptiness are obviously true for types with information θ''_i. For types with information θ'_i, we proceed by induction on k. The initial case ($k = 0$) is trivially true; for the inductive step, suppose that $k \geq 0$ is such that the claims hold, and show that this implies that it holds for $k+1$. For u.h.c., fix player i and take convergent sequence of types $(t_i^\nu)_{\nu \in \mathbb{N}}$ with limit t_i and strategy $s_i \in \bigcap_{\nu \in \mathbb{N}} RP_i^{k+1}(t_i^\nu)$. For each $\nu \in \mathbb{N}$ take conjecture μ_i^ν that justifies the inclusion of s_i in $RP_i^{k+1}(t_i^\nu)$. We know from compactness of $\Delta(S_j \times T_j \times \Omega)$ that there exists some convergent subsequence of $(\mu_i^\nu)_{\nu \in \mathbb{N}}$, $(\mu_i^{\nu_m})_{m \in \mathbb{N}}$, whose limit we denote by μ_i. Continuity of marginalization guarantees that μ_i is consistent with t_i, and by u.h.c. of best responses a_i is a best response to μ_i for type t_i. Under the induction hypothesis, RP_j^k is u.h.c., and hence RP_j^k is closed. It follows that $\mu_i[RP_j^k \times \Omega] \geq \mathrm{limsup}_{m \to \infty} \mu_i^{\nu_m}[RP_j^k \times \Omega] = 1$. This way, we conclude that $s_i \in RP_i^{k+1}(t_i)$, and thus, that RP_i^{k+1} is u.h.c. For nonemptiness of $RP_i^{k+1}(t_i)$ notice that we know that RP_j^k is nonempty-valued and hence there exist conjectures μ_i for t_i concentrated on RP_j^k. Set then $p := \mu_i[S_j \times T_j \times \{\omega^0\}]$ and $\eta_i[a_j] = \mu_i[T_j \times \{(a_j, \omega^0)\}]$ for all $a_j \in A_j$. Obviously, $\eta_i \in \Delta(\mathcal{B}_j)$. Hence, if the 'hybrid' best response to p and μ_i is unique, then it is in $\mathcal{B}_i$ and hence also in $RP_i^{k+1}(t_i)$. Otherwise, consider sequence of types $(t_i^\nu)_{\nu \in \mathbb{N}}$ such that $\tau_i(t_i^\nu) = (1 - 1/\nu) \cdot \tau_i(t_i) + (1/\nu) \cdot t_i^i$, where t_i^i is the type consistent with common belief in ω^i. Obviously, $(\tau_i^\nu)_{\nu \in \mathbb{N}}$ approaches t_i. Moreover, p^ν and η_i^ν are defined from t_i^ν analogously as p and η_i are for type t_i, and hence (using Assumption 1) for ν large enough the 'hybrid' best response is unique. Hence there exist $\bar{\nu}$ and a_i such that $s_i \in \bigcap_{\nu \geq \bar{\nu}} RP_i^{k+1}(t_i^\nu)$ and thus $s_i \in RP_i^{k+1}(t_i)$ from u.h.c. of RP_i^{k+1}.

A.3 Proof of Lemma 4

The proof of the Lemma exploits the following auxiliary solution concept: For each type t_i let $W_i^{\mathcal{C}}(t_i) := \bigcap_{k \geq 0} W_i^{\mathcal{C},k}(t_i)$, where $W_i^{\mathcal{C},0}(t_i) := \mathcal{C}_i(\hat{\theta}_i(t_i))$ and then, for every $k \geq 0$,

$$W_i^{\mathcal{C},k+1}(t_i) := \left\{ a_i \in \hat{A}_i : \begin{array}{ll} \multicolumn{2}{l}{\exists \mu_i \in \Delta(\hat{A}_j \times T_j \times \Omega) \text{ such that:}} \\ (i) & \mathrm{marg}_{T_j \times \Omega} \mu_i = \tau_i(t_i) \\ (ii) & (a_j, t_j, \omega) \in \mathrm{supp}\ \mu_i \implies a_j \in W_j^{\mathcal{C},k}(t_j) \\ (iii) & \underset{a'_i \in \hat{A}_i}{\mathrm{arg\,max}} \sum_{a_j \in \hat{A}_j} \sum_{\omega \in \Omega} \mu_i[T_j \times \{(a_j, \omega)\}] \cdot \hat{u}_i(a_i, a_j, \omega) = \{a_i\} \end{array} \right\}.$$

Lemma 7 *For every $k \geq 0$, every player i, every state ω and every action $a_i \in \mathcal{C}_i^k(\theta_i(\omega))$ there exists some finite type $t_i^{a_i,\omega} \in T_i^*$ with information $\theta_i(\omega)$ such that* $\text{ICR}_i^{k+1}(t_i^{a_i,\omega}) = \{a_i\}$.

Proof. We proceed by induction on k. The initial step ($k = 0$) holds by definition. For the inductive step, let $k \geq 0$ be such that the claim holds; we verify that it also holds for $k+1$. Fix player i and state ω. If $\mathcal{C}^{k+1}(\theta_i(\omega))$ is empty then the claim holds trivially. Otherwise, fix action $a_i \in \mathcal{C}_i^{k+1}(\theta_i(\omega))$ and conjecture μ_i which justifies the inclusion of a_i in $\mathcal{C}_i^{k+1}(\theta_i(\omega))$. We know from the inductive hypothesis that $\forall (a_j, \omega) \in \text{supp } \mu_i$, there exists some finite type $t_j^{a_j,\omega}$ with information $\theta_j(\omega)$ and for which $\text{ICR}_j^{k+1}(t_j^{a_j,\omega}) = \{a_j\}$. Define $t_i^{a_i,\omega}$ as the type with information θ_i and beliefs $\tau_i[E] := \mu_i\left[\left\{(a_j,\omega) \in \hat{A}_j \times \Omega : (t_j^{a_j,\omega},\omega) \in E\right\}\right]$ for every measurable $E \subseteq T_j$. Obviously, $t_i^{a_i,\omega}$ is well-defined and finite. Pick now arbitrary conjecture $\mu_i' \in C_i^{k+1}(t_i^{a_i,\omega})$, and notice that for every $(a_j,\omega') \in \text{supp } \mu_i$ we have that:

$$\begin{aligned}
\mu_i'\left[T_j \times \{(a_j,\omega')\}\right] &= \mu_i'\left[\left\{t_j^{a_j,\omega''} : \omega' \in \theta_j(\omega'') \cap \theta_i(\omega)\right\} \times \{(a_j,\omega')\}\right] \\
&= \mu_i'\left[\left\{t_j^{a_j,\omega''} : \omega' \in \theta_j(\omega'') \cap \theta_i(\omega)\right\} \times \hat{A}_j \times \{\omega'\}\right] \\
&= \tau_i\left[\left\{t_j^{a_j,\omega''} : \omega' \in \theta_j(\omega'') \cap \theta_i(\omega)\right\} \times \{\omega'\}\right] = \mu_i[(a_j,\omega')]
\end{aligned}$$

Clearly, it follows that $\text{ICR}_i^{k+2}(t_i^{a_i}) = \{a_i\}$. ■

Lemma 8 *For every i, every finite type $t_i \in T_i^*$ and every $a_i \in \text{ICR}_i^{\mathcal{C}}(t_i)$ there exists a sequence of finite types $(t_i^\nu)_{\nu\in\mathbb{N}}$ in T_i^* converging to t_i such that $s_i \in W_i^{\mathcal{C}}(t_i^\nu)$ for every $\nu \in \mathbb{N}$.*

Proof. Fix arbitrary finite type space $(T_i, \hat{\theta}_i, \tau_i)_{i=1,2}$. Then, for every $\nu \in \mathbb{N}$ define type space $(T_i^\nu, \hat{\theta}_i^\nu, \tau_i^\nu)_{i=1,2}$ by setting for each player i:

- Set of types $T_i^\nu := \{\nu\} \times \left\{(a_i,t_i),(a_i,t_i^{a_i}) : t_i \in T_i \text{ and } a_i \in \text{ICR}_i^{\mathcal{A}}(t_i)\right\}$, where $t_i^{a_i}$ is constructed as in Lemma 7. Obviously, T_i^ν is a finite set.
- Information-map $\hat{\theta}_i^\nu : T_i^\nu \to \Theta_i$ given by $(\nu, a_i, t_i) \mapsto \hat{\theta}_i(t_i)$.
- Finally, in order to define belief-maps, for state ω and action $a_i \in \mathcal{C}_i(\theta_i(\omega))$ let $\mu_i^{a_i,\omega}$ denote a conjecture over $\hat{A}_j \times \Omega$ that justifies the inclusion of a_i in $\mathcal{C}_i(\theta_i(\omega))$. Then, define player i's belief-map $\tau_i^\nu : T_i^\nu \to \Delta(T_j^\nu \times \Omega)$ as follows:

$$(\nu, a_i, t_i) \mapsto \tau_i^\nu(\nu,a_i,t_i)[(\nu,a_j,t_j,\omega')] := \begin{cases} \left(1-\frac{1}{\nu}\right)\tau_i(t_i)\,[t_j] & \text{if } t_j \in T_j, \\ \left(\frac{1}{\nu}\right) 1_{\{t_j^{a_j,\omega'}\}}(t_j)\cdot \mu_i^{a_i,\omega}[(a_j,\omega')] & \text{otherwise,} \end{cases}$$

  for every $(\nu, a_j, t_j, \omega') \in T_j^\nu \times \Omega$ such that (t_j, ω') is in the support of $\mu_i^{a_i,\omega}$, and $t_j^{a_j,\omega'}$ is constructed as in Lemma 7. The finiteness of the set of types guarantees that these belief-maps are well-defined and continuous, and that every type in T_i^ν and T_j^ν is finite.

We claim that the following hold: (i) $\forall t_i \in T_i$, each $((\nu,a_i,t_i))_{\nu\in\mathbb{N}}$ converges to t_i; (ii) $\forall t_i \in T_i$ and $\forall a_i \in \text{ICR}_i^{\mathcal{C}}(t_i)$, $a_i \in W_i^{\mathcal{C}}(\nu,a_i,t_i)$ for every $\nu \in \mathbb{N}$. To prove the claim of the lemma, fix player i and pick arbitrary finite type $\bar{t}_i \in T_i^*$ and action $\bar{a} \in \text{ICR}_i^{\mathcal{C}}(\bar{t}_i)$. Since $\bar{t}_i$ is finite we know that there exists some finite type space $(T_i, \hat{\theta}_i, \tau_i)_{i=1,2}$ where T_i includes some type $\hat{t}_i$ that induces $\bar{t}_i$. Consider the sequence of finite type spaces $((T_i^\nu, \hat{\theta}_i^\nu, \tau_i^\nu)_{i=1,2})_{\nu \in \mathbb{N}}$ constructed above. Byf type-space invariance, $a_i \in \text{ICR}_i^{\mathcal{C}}(\hat{t}_i)$ and by the construction above we know that $\forall \nu \in \mathbb{N}$ there exists some type $t_i^\nu \in T_i^\nu$ such that $\bar{a}_i \in W_i^{\mathcal{C}}(t_i^\nu)$. Let $(\bar{t}_i^\nu)_{\nu \in \mathbb{N}}$ the sequence in the universal type space induced by $(t_i^\nu)_{\nu \in \mathbb{N}}$. Again, because of type-space invariance we know that $\bar{a}_i \in W_i^{\mathcal{C}}(\bar{t}_i^\nu)$ for every $\nu \in \mathbb{N}$.[24] Finally, since we know that $(t_i^\nu)_{\nu \in \mathbb{N}}$ converges to $\hat{t}_i$ we also know that $(\bar{t}_i^\nu)_{\nu \in \mathbb{N}}$ converges to $\bar{t}_i$ and hence, the proof is complete. ■

For the following lemma let $m \in \mathbb{N}$ be such that $\mathcal{C}_i(\theta_i) = \mathcal{C}_i^m(\theta_i)$ for every player i and information type θ_i. Then, we have that:

Lemma 9 *For every $k \geq 1$, every player i, every finite type $t_i \in T_i^*$ and every action $a_i \in W_i^{\mathcal{C},k}(t_i)$ there exists some finite type $t_i^k \in T_i^*$ such that: $\hat{\theta}_i(t_i^k) = \hat{\theta}_i(t_i)$, $\pi_i^k(t_i^k) = \pi_i^k(t_i)$, and $\text{ICR}_i^{m+k+2}(t_i^k) = \{a_i\}$.*

Proof. We proceed by induction on k: *Initial step* ($k = 1$). Set $\ell = 1$. Fix player i, finite type $\bar{t}_i$, action $\bar{a}_i \in W_i^{\mathcal{C},\ell}(\bar{t}_i)$ and conjecture $\bar{\mu}_i$ that justifies the inclusion of $\bar{a}_i$ in $W_i^{\mathcal{C},\ell}(\bar{t}_i)$. Then, we know by Lemma 7 that $\forall (a_j, t_j) \in \text{supp}(\text{marg}_{\hat{A}_j \times T_j} \bar{\mu}_i)$, there exists a finite type $t_j^{\ell-1}(a_j, t_j)$ with the same information as t_j and s.t. $\text{ICR}_j^{m+\ell}(t_j^{\ell-1}(a_j, t_j)) = \{a_j\}$. Then, let type t_i^ℓ have information $\hat{\theta}_i(\bar{t}_i)$ and beliefs $\tau_i^\ell[E] := \bar{\mu}_i\left[\left\{(a_j, t_j, \omega) \in \hat{A}_j \times T_j \times \Omega : (t_j^{\ell-1}(a_j, t_j), \omega) \in E\right\}\right]$, for every measurable $E \subseteq T_j$. Obviously, t_i^ℓ is well-defined and finite, and has the same ℓ^{th}-order beliefs as $\bar{t}_i$—and thus, as $\hat{t}_i$. Finally, pick arbitrary conjecture μ_i inducing t_i^ℓ and puts probability 1 on the graph of $\text{ICR}_j^{m+\ell}$ and notice that for every (a_j, ω) we have that:

$$
\begin{aligned}
&\mu_i[T_j \times \{(a_j, \omega)\}] = \\
&\quad = \mu_i\left[\left\{t_j^{\ell-1}(a_j', t_j')) : (a_j', t_j') \in \hat{A}_j \times T_j, a_j \in \text{ICR}_j^{m+\ell}(t_j^{\ell-1}(a_j', t_j'))\right\} \times \{(a_j, \omega)\}\right] \\
&\quad = \mu_i\left[\hat{A}_j \times \left\{t_j^{\ell-1}(a_j', t_j') : (a_j', t_j') \in \hat{A}_j \times T_j, \text{ICR}_j^{m+\ell}(t_j^{\ell-1}(a_j', t_j')) = \{a_j\}\right\} \times \{\omega\}\right] \\
&\quad = \tau_i^\ell\left[\left\{t_j^{\ell-1}(a_j', t_j') : (a_j', t_j') \in \hat{A}_j \times T_j, \text{ICR}_j^{m+\ell}(t_j^{\ell-1}(a_j', t_j')) = \{a_j\}\right\} \times \{\omega\}\right] \\
&\quad = \bar{\mu}_i\left[\left\{(a_j', t_j') \in \hat{A}_j \times T_j : \text{ICR}_j^{m+\ell}(t_j^{\ell-1}(a_j', t_j')) = \{a_j\}\right\} \times \{\omega\}\right] \\
&\quad = \bar{\mu}_i[T_j \times \{(a_j, \omega)\}].
\end{aligned}
$$

Clearly, it follows that $\text{ICR}_i^{m+\ell+1}(t_i^\ell) = \{\bar{a}_i\}$.

Induction step. Suppose that $k \geq 1$ is such that the claim holds. Then, to verify that it also does for $k+1$ simply repeat, verbatim, the steps of the initial step by replacing index $\ell = 1$ by $\ell = k+1$ and noticing that the existence of types $t_j^{\ell-1}(\,\cdot\,)$ is not due to Lemma 7, but due to the induction hypothesis, instead. ■

Proof of Lemma 4: Fix finite type $t_i \in T_i^*$ and action $a_i \in \text{ICR}_i^{\mathcal{C}}(t_i)$. Then, we know from Lemma 8 that there exists a sequence of finite types $(\hat{t}_i^\nu)_{\nu \in \mathbb{N}}$ in T_i^* converging to t_i such that

[24] The proofs that $\text{ICR}_i^{\mathcal{C}}$ and $W_i^{\mathcal{C}}$ are type-space invariant is analogous to that of Part (i) of Theorem 1: simply substitute RP_i^k for $\text{ICR}_i^{\mathcal{C},k}$ in one case, and for $W_j^{\mathcal{C},k}$ in the other.

$a_i \in W_i^{\mathcal{C}}(\hat{t}_i^\nu)$ for every $\nu \in \mathbb{N}$. It follows from Lemma 9 that $\forall \nu \in \mathbb{N}$ there exists a sequence of finite types $(t_i^{\nu,k})_{k\in\mathbb{N}}$ in T_i^* converging to $\hat{t}_i^\nu$ such that $\text{ICR}_i(t_i^{\nu,k}) = \{a_i\}$ for all $k \in \mathbb{N}$. Thus, if for each $\nu \in \mathbb{N}$ we set $t_i^\nu = t_i^{\nu,k}$, $(t_i^\nu)_{\nu\in\mathbb{N}}$ is a sequence of finite types in T_i^* converging to t_i such that $\text{ICR}_i(t_i^\nu) = \{a_i\}$ for every $\nu \in \mathbb{N}$.

A.4 Proof of Part (iii)

Proof of Lemma 3. The claim is obvious for types with information θ_i'': for these types a_i^* is weakly dominant in (G^*, Ω) and R_i removes all weakly dominated strategies (of types with information θ_i'') in the first round, and $\hat{a}_i$ is strictly dominant in $\hat{G}$ and ICR_i removes all strictly dominated actions in the first round. Thus, we only need to complete the proof for types with information θ_i'. We proceed by induction. The initial case $(k = 0)$ is trivially true, so we can focus on the inductive step. Suppose that $k \geq 0$ is such that the claim holds; we verify next that so does it for $k+1$. To see it fix type $t_i \in T_i^*$ with information θ_i' and define map $\beta_i : M_i(t_i) \to N_i(t_i)$ where $M_i(t_i) := \Big\{\mu_i \in C_i(t_i) : \mu_i[\{(a_j^*, \omega^i)\} \times T_j] = \mu_i[S_j \times \{\omega^i\} \times T_j]\Big\}$, $N_i(t_i) := \Big\{\mu_i \in \Delta(\hat{A}_j \times T_j \times \Omega) : \text{marg}_{T_j\times\Omega}\mu_i = \tau_i(t_i)\Big\}$, by setting:

$$\begin{aligned}\beta_i(\mu_i)[E] = \mu_i\left[\{(a_j, t_j, \omega) \in A_j \times T_j \times \Omega : (a_j, t_j, \omega) \in E\}\right] \\ + \mu_i\left[\{(a_j, t_j, \omega) \in \{a_i^*\} \times T_j \times \Omega : (\hat{a}_j, t_{j,\omega}) \in E\}\right]\end{aligned}$$

for all $\mu_i \in M_i(t_i)$. β_i is a well-defined bijection that satisfies: (i) a_i is a best response to μ_i in (G^*, Ω) if and only it is a best response to $\beta_i(\mu_i)$ in $\hat{G}$; (ii) μ_i puts probability 1 on the graph of R_j^k if and only if $\beta_i(\mu_i)$ puts probability 1 on the graph of ICR_j^k. Hence, μ_i justifies the inclusion of strategy $s_i = a_i$ in $R_i^{k+1}(t_i)$ if and only if conjecture $\beta_i(\mu_i)$ justifies the inclusion of action a_i in $\text{ICR}_i^{k+1}(t_i)$. The result for the $T_i^\dagger$ space follows from the same logic. ■

Lemma 10 *For every player i: (i) with space of uncertainty* Ω*,* $\mathcal{C}_i(\theta_i^{'}) = \mathcal{B}_i$ *and* $\mathcal{C}_i(\theta_i^{''}) = \{\hat{a}_i\}$*; (ii) with space of uncertainty* $\Omega^\dagger$*,* $\mathcal{C}_i(\theta_i^{'}) = \mathcal{B}_i^\dagger$ *and* $\mathcal{C}_2(\theta_2^{''}) = \{\hat{a}_2\}$*.*

Proof. We complete the proof for part (i) (for part (ii), it suffices to substitute $\mathcal{B}$ and $\mathcal{C}$ for $\mathcal{B}^\dagger$ and $\mathcal{C}^\dagger$, respectively). That $\mathcal{C}_i^k(\theta_i'') = \{\hat{a}_i\}$ for all $k \geq 0$ is immediate. For information types θ_i' we show first that $\mathcal{B}_i^k \subseteq C_i^k(\theta_i')$ for every $k \geq 0$. For the initial case $(k = 0)$ simply notice that for t_i that has common belief in ω^i, $\text{ICR}_i^2(t_i) = \{a_i^i\}$ and hence $a_i^i \in C_i^0(\theta_i')$. Suppose now that, by induction, $k \geq 0$ is such that the claim holds, and pick $a_i \in \mathcal{B}_i^{k+1}$, $p \in [0,1]$, and $\mu_i \in \Delta(\mathcal{B}_j^k)$ s.t. a_i maximizes the expected payoff induced by p and μ_i. Set

$$\eta_i[(a_j, \omega)] := \begin{cases} p \cdot \mu_i[a_j] & \text{if } \omega \neq \omega^i \text{ and } a_j \in A_j, \\ 1 - p & \text{if } \omega = \omega^i \text{ and } a_j = \hat{a}_j, \end{cases}$$

for every $(a_j, \omega) \in \hat{A}_j \times \Omega$. Obviously, η_i satisfies conditions (i) and (ii) w.r.t. a_i in the definition of $\mathcal{C}_i^{k+1}(\theta_i')$. Finally, we show that $C_i^k(\theta_i') \subseteq \mathcal{B}_i$ for every $k \geq 0$. For the initial case $(k = 0)$ pick $a_i \in C_i^0(\theta_i')$ and type t_i such that $\text{ICR}_i(t_i) = \{a_i\}$. Then, we know from Lemma 3 and nonemptiness of RP_i that $RP_i(t_i) = \{a_i\}$. Thus, it follows that $a_i \in \mathcal{B}_i$. Now, we know from Lemma 4 that $\mathcal{C}_i^0(\theta_i') = \mathcal{C}_i(\theta_i')$ and hence, the proof is complete. ■

Proof of Lemma 5. The proof is analogous to that of Lemma 3, by induction. For the initial step take Lemma 10 into account, and for the inductive step, notice that because of the induction hypothesis, map β_i, as defined in the proof of the Lemma 3, μ_i puts probability 1 on the graph of RP_j^k if and only if $\beta_i(\mu_i)$ puts probability 1 on the graph of $\text{ICR}_j^{\mathcal{C},k}$. ■

Proof of Lemma 6. Again, the proof is analogous to that of Lemma 3, by induction. For the initial step, take Lemma 10 into account and for the inductive step, by the induction hypothesis, map β_i, as defined in the proof of the Lemma 3, μ_i puts probability 1 on the graph of $RP_j^{\dagger,k}$ if and only if $\beta_i(\mu_i)$ puts probability 1 on the graph of $\text{ICR}_j^{\mathcal{C}^\dagger,k}$. ■

Proof of Part (iii): For Theorem 1 fix player i, finite type $t_i \in T_i^*$ and strategy $s_i \in RP_i(t_i)$. If $\hat{\theta}_i(t_i) = \theta_i''$ then the claim is trivially true. If $\hat{\theta}_i(t_i) = \theta_i'$, we know that $s_i = a_i$ for some action $a_i \in A_i$ and thus, it follows from Lemma 5 that $a_i \in \text{ICR}_i^{\mathcal{C}}(t_i)$. By Lemma 4, there exists a sequence of finite types $(t_i^\nu)_{\nu\in\mathbb{N}}$ converging to t_i such that $\text{ICR}_i(t_i^\nu) = \{a_i\}$ for every $\nu \in \mathbb{N}$. By applying Lemma 3 we conclude that, indeed, $R_i(t_i^\nu) = \{s_i\}$ for every $\nu \in \mathbb{N}$. For Theorem 2, simply repeat the argument substituting T_i^*, $\mathcal{C}$, ICR and R_i with $T_i^\dagger$, $\mathcal{C}^\dagger$, $\text{ICR}^\dagger$ and $R_i^\dagger$, respectively, and applying lemma 6 instead of lemma 5.

B Other Results

Proof of Corollary 2. Let $F_i : T_i^* \rightrightarrows A_i$ be s.t. $F_i(t_i)$ denotes the set of actions $a_i \in R_i(t_i)$ s.t. for any neighborhood $N \in \mathcal{N}(t_i)$ of t_i, there exists an open subset $U \subset \mathcal{N}(t_i)$ s.t. $a_i \in R_i(t_i')$ $\forall t_i' \in U$. Notice first that F_i is u.h.c.. To see this, proceed by contradiction and suppose that $(t_i^\nu)_{\nu\in\mathbb{N}}$ converges to t_i, $a_i \in F_i(t_i^\nu)$ for every $\nu \in \mathbb{N}$ and $a_i \notin F_i(t_i)$. By, u.h.c. of R_i we have $a_i \in R_i(t_i)$. Then there exists $N \in \mathcal{N}(t_i)$ s.t. $\forall V \subseteq N$ there is some $t_i' \in V$ s.t. $a_i \notin R_i(t_i')$. But this is a contradiction: $N \in \mathcal{N}(t_i^n)$ for large enough ν and $a_i \in F_i(t_i^\nu)$.

To see that $F_i(t_i) \subseteq RP_i(t_i)$, pick an arbitrary $a_i \in F_i(t_i)$ and $N \in \mathcal{N}(t_i)$. By Theorem 1, there exists an open and dense $X \subseteq T_i^*$ in which R_i and RP_i coincide. Then, there exists some open $U \subseteq N$ s.t. $a_i \in R_i(t_i') = RP_i(t_i')$ for every $t_i' \in U \cap X \subseteq N$. Hence, $\forall N \in \mathcal{N}(t_i)$ there exists t_i^N s.t. $a_i \in RP_i(t_i^N)$. Since RP_i is u.h.c., we have $a_i \in RP_i(t_i)$.

For the other inclusion, pick $a_i \in RP_i(t_i)$. If t_i is finite pick sequence $(t_i^\nu)_{\nu\in\mathbb{N}}$ converging to t_i s.t. $RP_i(t_i^\nu) = \{a_i\}$ $\forall \nu \in \mathbb{N}$. Obviously, $a_i \in R_i(t_i)$. In addition, u.h.c. of RP_i implies that $\forall \nu \in \mathbb{N}$ there exists some open $U^\nu \in \mathcal{N}(t_i^\nu)$ s.t. $RP_i(t_i') = \{a_i\}$ $\forall t_i'^\nu$. Since $\forall N \in \mathcal{N}$ there exists some $\nu \in \mathbb{N}$ s.t. $t^\nu \in N$, there also exists some $U \subseteq N$, $U = U^\nu \cap N$, s.t. $a_i \in RP_i(t_i') \subseteq R_i(t_i')$ $\forall t_i' \in U$. That is, $a_i \in F_i(t_i)$. Finally, the u.h.c. of F_i implies that the inclusion is also true for nonfinite types. ■

Proof of Proposition 1. 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 $t_i \in \check{T}_i$. Since $a_i^i = a_i^j = a_i^*$, it follows from Assumption 1 that $a_i^j = a_j^i$, and hence, that $\mathcal{B}_i = \{a_i^*\}$, which in turn implies $R_i(t_i) = RP_i(t_i) = \{a_i^*\}$ for any $t_i \in \check{T}_i$. R_i's u.h.c. then implies 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 an open and dense subset of T_i^* and such that $R_i(t_i) = \{a_i^*\}$ for every $t_i \in T_i^*$. ■

Proof of Proposition 2. Under the assumptions of the proposition, w.l.o.g. let $u_i^*(a) = 0$

for any non-Nash profile a. Then, note that for any i, $p \in [0,1]$ and $a_i \neq a_i^i, a_i^j$, 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:

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

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

Proof of Proposition 3. Fix player i. By Theorem 2, there exists some dense subset $\check{T}_i \subseteq T_i^\dagger$ s.t. $|R_i(t_i)| = 1$ and $R_i(t_i) = RP_i^\dagger(t_i)$ $\forall t_i \in \check{T}_i$. Since a^1 is a Nash equilibrium, by Assumption 1 $\mathcal{B}_1^\dagger = \{a_1^1\}$ and $\mathcal{B}_2^\dagger = \{a_2^1\}$. The rest of the proof is the same as in the proof of Proposition 1, replacing $\mathcal{B}$ with $\mathcal{A}$ and RP with $RP^\dagger$. ■

References

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

2. Alaoui, L. and A. Penta (2019) "Coordination and Sophistication," *Barcelona GSE w.p.*.

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

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

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

6. 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.

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

8. Bergemann, D. and S. Morris (2005) "Robust Mechanism Design," *Econometrica*, 1521–34.

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

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

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

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

13. Binmore, K. (1987-88) "Modeling Rational Players," *Econ. and Phil.*, 3:179–214; 4:9–55.

14. Brandenburger, A. and E. Dekel (1993) "Hierarchies of Beliefs and Common Knowledge," *Journal of Economic Theory*, 59, 189–198.

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

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

17. Chen, Y., (2012) "A Structure Theorem for Rationalizability in the Normal Form of Dynamic Games," *Games and Economic Behavior*, 75, 587-597.

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

19. 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.

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

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

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

23. Doval, L. and J. Ely (2016) "Sequential Information Design," *mimeo.*

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

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

26. 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.

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

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

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

30. Kalai, E. (2004) "Large Robust Games," *Econometrica*, 72(6), 1631–1665.

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

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

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

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

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

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

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

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

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

40. Ollar, M. and A. Penta (2019) "Implementation via Transfers with Identical but Unknown Distributions," *Barcelona GSE working paper*.

41. Oury, M. and O. Tercieux (2012) "Continuous Implementation," *Econometrica*, 1605–37.

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

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

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

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

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

47. Robson, A. J. (1994) "An "Informationally Robust Equilibrium" for Two-Person Nonzero-Sum Games," *Games and Economic Behavior*, 7, 133-245.

48. Reny, P. J. and A. J. Robson (2004) "Reinterpreting mixed strategy equilibria: a unification of the classical and Bayesian Views," *Games and Economic Behavior*, 48, 355-384.

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

50. Samuelson, L. (1998) *Evolutionary Games and Equilibrium Selection*, MIT Press.

51. Schelling, T.C. (1960) *The Strategy of Conflict*, Harvard Univ. Press.

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

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

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

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

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

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