HIGHER SCHOOL OF ECONOMICS
NATIONAL RESEARCH UNIVERSITY

Antonio Penta, Peio Zuazo-Garin

RATIONALIZABILITY, OBSERVABILITY AND COMMON KNOWLEDGE

Working Paper WP9/2021/05
Series WP9
Research of economics and finance

Moscow
2021

УДК 330.4
ББК 65.01
Р41

Editor of the Series WP9
"Research of economics and finance"
Maxim Nikitin

Penta, A., Zuazo-Garin, P.
P41 Rationalizability, Observability and Common Knowledge* [Electronic resource] : Working paper WP9/2021/05 / A. Penta, P. Zuazo-Garin ; National Research University Higher School of Economics. – Electronic text data (480 Kb). – Moscow : Higher School of Economics Publ. House, 2021. – (Series WP9 "Research of economics and finance"). – 55 p.

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 captures 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 (*eductive coordination*); (ii) in a class of games which also includes other well-known families of coordination games, RCBR generically selects components of the Stackelberg profiles (*Stackelberg selection*); (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 all of the above RCBR generically selects the equilibrium of the static game most favorable to player 1 (*pervasiveness of first-mover advantage*).

УДК 330.4
ББК 65.01

Keywords: eductive coordination; extensive form uncertainty; first-mover advantage; Kreps Hypothesis; higher-order beliefs; rationalizability; robustness; Stackelberg selections

Penta Antonio, ICREA, Universitat Pompeu Fabra and Barcelona-GSE, Spain; E-mail: antonio.penta@upf.edu

Zuazo-Garin Peio, ICEF, National Research University Higher School of Economics, Russian Federation; E-mail: p.zuazogarin@hse.ru

* We dedicate this article to the memory of Bill Sandholm, who gifted us with invaluable comments and encouragement in the early stages of this project. Bill's acumen, depth and generosity are sorely missed; his friendship, is irreplaceable. We also thank Pierpaolo Battigalli, Eddie Dekel, Laura Doval, Glenn Ellison, Drew Fudenberg, Mike Peters, Lones Smith, Stephen Morris, Jakub Steiner, and Muhamet Yildiz, as well as seminar audiences at Cambridge, UW-Madison, UCSD, UPV/EHU, Univ. Pu ́blica de Navarra, PSE, CERGE-EI, UPF, TSE, Groningen, Bonn, EUI, Alicante, Korea Univ., Kyung Hee Univ., and participants of the 2017 SISL Summer Conference (Caltech), 2017 HKU-UBC Summer Workshop in Economic Theory, 2017-StratEmon Workshop (Bocconi), 2018 Southampton-Bristol Economic Theory Workshop, 2018 Warwick Economic Theory Workshop, 2018 LOFT Conference (Bocconi). This research benefited from the support of the Spanish Ministry of Science and Innovation, through the Severo Ochoa Programme for Centres of Excellence in R&D (CEX2019- 000915-S). Antonio Penta acknowledges the financial support of the ERC Starting Grant #759424.

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), Lipman (2003), Weinstein and Yildiz (2007, 2011, 2013, 2016), etc.). In contrast, the assumption of common knowledge of the extensive form has hardly been challenged.[1] Yet, uncertainty over the extensive form is key to many strategic situations. It is clearly paramount in military applications, but economic settings abound in which players are uncertain over the moves that are available to their opponents, or about their opponents' information about their moves, etc., in ways which need not match the typical common knowledge assumptions that are implicit in standard economic 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

[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)), 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 5.

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 nonempty refinement which characterizes the *regular predictions* of RCBR, i.e., those which do not depend on knife-edge, nongeneric restrictions on the belief hierarchies. We then explore the implications of these results in classes of games in which they are especially sharp or significant, and show that they provide theoretical foundations to intuitive yet hard to explain predictions in disparate classes of games.

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 ex-

[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.

tensive 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 be achieved on the basis of purely 'eductive' mechanisms (Binmore (1987-88)), in the absence of focal points and with no 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. 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

[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 'evolutionary approach', in which solution concepts are interpreted as the steady state of an underlying learning or evolutionary 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).

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, to 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, *timing* of moves alone (plus irreversibility of choices) may determine the attribution of the strategic advantage, independently of the actual observability of actions. This message is clearly at odds with the received game theoretic wisdom that observability, not timing, is key to ensure the upper hand in a strategic situation. Our results show that this classical insight is somewhat fragile, and in fact overturned, when one considers even arbitrarily small departures from the standard assumptions of common knowledge on the extensive form.

A large experimental literature has explored the impact of timing on individuals' choices in a static game, with findings that are often difficult to reconcile with the received game theoretic wisdom. For instance, it is well-known (see, e.g., Camerer (2003)) that asynchronous moves in the Battle of the Sexes systematically select the Nash equilibrium most favorable to the first mover, 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.

The rest of the paper is organized as follows: Section 1.1 presents an illustrative example; Section 2 introduces the model; Section 3 contains the general characterization of the R correspondence, and Section 4 explores some of its implications for eductive coordination and robust refinements, as well as the variations with one-sided uncertainty. Section 5 discusses the most closely related literature, and in particular the connections with the closely related work by Weinstein and Yildiz (2007). Section 6 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 is common knowledge that the game is static, everything is rationalizable (and, thus, consistent with RCBR).

In an influential paper, which will be further discussed below, Weinstein and Yildiz (2007) characterize the set of predictions which can be made for static games like this, that would retain their validity under small perturbations of common knowledge assumptions on players' payoffs. Their results imply that no outcome can be robustly ruled-out in this game, under their form of perturbations. Here we consider different perturbations of the common knowledge assumptions: namely, we maintain that payoffs are common knowledge, but we introduce higher-order uncertainty over the observability of actions. As we will show, this change has profound effects on the insights that emerge from the analysis.

For instance, suppose that players commonly agree that player 1 chooses before 2, but there is uncertainty over whether his action will be observed. Let ω^0 denote the state of the world in which actions are not observable, and 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 is common belief that the state is ω^1. Then, 2 expects 1 to choose U, and hence L is his 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 reply 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 she is merely believed to have it at some arbitrarily high order of beliefs. Proposition 3 in Section 4 implies that, if the only uncertainty concerns the observability of 1's action, then this selection actually 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 4 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 games, 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 4.

As a comparison, Weinstein and Yildiz (2007, WY) maintain common knowledge that the game is static, and consider perturbations of belief hierarchies over a 'rich' space of payoff uncertainty, which contains strict dominance states for every player's action. Hence, any action profile could be used to start the 'infection' in the argument above. Thus, because of their richness assumption, in WY's space there would be belief hierarchies similar to the ones above, in which higher-order beliefs also trace back to profile (D, L), which would thus be uniquely selected along the sequence. This implies that no refinement of rationalizability is robust in their setting, and hence – under their perturbations of common knowledge of payoffs – no outcome can be robustly ruled out in the game above. The qualitative message that emerges from our paper is thus very different from WY's *unrefinability* result.

As shown by this example, one key difference between our analysis and WY's is due to the fact that, given the nature of extensive-form uncertainty that we consider, only the two backward induction outcomes (or just (U, L), in the first case) could be used to ignite the 'infection argument' in our setting. But this is only *one* of the points of departure from WY, and it doesn't suffice to explain the difference in the general results. Because of the particular configuration of payoffs, the infection argument in this example only involved a standard chain of (static) best responses. In general games, however, the robust predictions also depend on the behavior of types who are uncertain over whether the game is static or dynamic, whose optimization problem therefore is a hybrid of the standard static and dynamic ones. These hybrid best responses carry over to the higher-order beliefs, and hence the way the infection spreads from one type to another will differ from WY's, and so will do the robust predictions. These further differences will be explained in Sections 3 and 5.

2 Model

2.1 Environment

Let $G^* := (A_i, u_i^*)_{i=1,2}$ denote a static two-player game, where for any $i = 1, 2$ A_i denotes i's set of actions, and and $u_i^* : A_1 \times A_2 \to \mathbb{R}$ his payoff function, all assumed common knowledge. We also let $A := A_1 \times A_2$. Similar to the example in Section 1.1, we introduce extensive-form uncertainty by letting $\Omega := \left\{\omega^0, \omega^1, \omega^2\right\}$ denote the set of states of the world: 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 3.) We maintain throughout the following assumption on G^*:

Assumption 1 *For each $i \in \{1,2\}$, $j \neq i$, and $a_i \in A_i$, $\exists!\ a_j^*(a_i) : \arg\max_{a_j \in A_j} u_j^*(a_j, a_i) = \{a_j^*(a_i)\}$ and for each $a_i, a_i' \in A_i$ s.t. $a_i \neq a_i'$, $u_i^*(a_i, a_j^*((a_i)) \neq u_i^*(a_i', a_j^*(a_i'))$.*

In words, the first part says that for each of player i's actions, j has a unique best response; the second part says that no two distinct actions of player i, when combined with the corresponding best replies of player j, yield the same the payoff to player i. This assumption, which is weaker than requiring that payoffs in G^* are in generic position, ensures that backward induction is well-defined, and identifies a unique outcome, 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, we will denote by $a^i = \left(a_1^i, a_2^i\right)$ the backward induction 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: There are two possible pieces of 'hard information' for a player: either 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 $\theta_i' := \left\{\omega^0, \omega^i\right\}$ and $\theta_i'' := \left\{\omega^j\right\}$. 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 $\hat{\theta}_i : T_i \to \Theta_i$

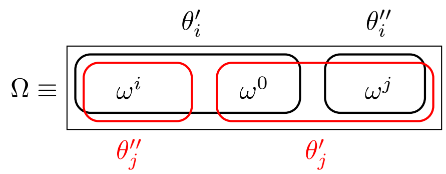

Figure 1: Information partitions on Ω.

is a Borel-measurable map that 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., $\tau_i(t_i)[\{(t_j,\omega) : \omega \in \hat{\theta}_i(t_i) \cap \hat{\theta}_j(t_j)\}] = 1$).

Each type t_i encodes a *belief hierarchy* about the states of the world, that consists of a *first-order* belief about Ω, a *second-order* belief about Ω and the opponents' first-order beliefs, and so on. Type t_i's first-order belief is obtained by taking the marginal of $\tau_i(t_i)$ over Ω, so as to obtain an element in $Z_1^1 := \Delta(\Omega)$; higher order beliefs are obtained following a customary recursive marginalization procedure, where type t_i's k-th order belief is an element of $Z_i^k = Z_i^{k-1} \times \Delta(Z_j^{k-1})$, and hence it consists of t_i's $(k-1)$-th order beliefs, plus his belief about player j's $(k-1)$-th order beliefs (see Appendix A for details). We let $\hat{\pi}_i^k(t_i)$ denote t_i's k-th order belief and $\hat{\pi}_i(t_i) := (\hat{\pi}_i^k(t_i))_{k\in\mathbb{N}}$, its complete belief hierarchy.

As shown by Mertens and Zamir (1985), it is possible to construct a *universal type space* $\mathcal{T}^* = (T_i^*, \hat{\theta}_i^*, \tau_i^*)_{i=1,2}$, in which the set of types T_i^* coincide with the set of all (mutually consistent) information-belief hierarchy pairs, (θ_i, π_i), endowed with the product topology. Hence, by construction, the universal type space is such that each type $t_i = (\theta_i, \pi_i)$ is such that $\theta_i = \hat{\theta}_i^*(t_i)$ and $\pi_i = \hat{\pi}_i^*(t_i)$ (see Appendix A for the detailed construction). A Cartesian subset $T_1' \times T_2' \subseteq T_1^* \times T_2^*$ is said to be *belief-closed* if, for each player i and each type $t_i \in T_i'$, the support of $\tau_i(t_i)$ is contained in $T_j' \times \Omega$; a type t_i is said to be *finite* if it is contained in some finite belief-closed subset.

Finally, for each $\omega \in \Omega$, we let $t_i^{CB}(\omega)$ denote the type corresponding to common belief in ω, i.e, that type with first order belief assigning probability 1 to ω, assigning probability 1 to the opponents assigning probability 1 to ω, and so on.

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:[4]

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

2.2 Solution Concept

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 $\hat{\theta}_i(t_i) = \theta_i^{''}$. Noting that, under Assumption 1, $a_i^*(\cdot)$ is the only sequentially rational strategy for all types who move second, RCBR can be captured by a tractable iterated deletion procedure in the interim strategic form. Specifically: for types such that $\hat{\theta}_i(t_i) = \theta_i^{''}$, the procedure uniquely

[4]We note that the dynamic games in states ω^1 and ω^2 have different normal forms, both of which are distinct from the one that corresponds to ω^0, G^*. Hence, our exercise cannot be described in terms of uncertainty over the extensive forms that may generate a given (common) strategic form.

selects $a_i^*(\cdot)$ from the first round on; for all other types, it consists of a standard iterated deletion of never best replies.

Formally: fix a type space $\mathcal{T} = (T_i, \hat{\theta}_i, \tau_i)_{i=1,2}$; for any i and t_i, let $R_i^0(t_i) := S_i(t_i)$. Then, recursively for $k = 1, 2, ...$, and letting $R_j^{k-1} := \{(s_j, t_j) : s_j \in R_j^{k-1}(t_j)\}$, we define $R_i^k(t_i)$ to be such that, for any $t_i \in T_i$: if $\hat{\theta}(t_i) = \theta_i''$, then $R_i^k(t_i) := \{a_i^*(\cdot)\}$; otherwise,

$$R_i^k(t_i) := \left\{ s_i \in R_i^{k-1}(t_i) : \begin{array}{c} \exists \mu_i \in \Delta(R_j^{k-1} \times \Omega) \text{ such that:} \\ (i) \quad \text{marg}_{T_j \times \Omega} \mu_i = \tau_i(t_i), \\ (ii) \quad s_i \in \arg\max_{s_i' \in S_i(t_i)} \sum_{\omega \in \Omega} \sum_{s_j \in S_j(\omega)} \mu_i[\{(s_j, \omega)\} \times T_j] \, u_i(s_i', s_j, \omega) \end{array} \right\}.$$

Finally, we let $R_i(t_i) := \bigcap_{k \geq 0} R_i^k(t_i)$.

It can be shown that, under Assumption 1, this solution concept is equivalent to 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 types such that (based on a standard duality argument, the round of weak dominance serves to capture *sequential* rationality). Hence, R is effectively 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 can also be 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.[5]

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

Example 1 Consider a type space $T_i := \left\{t_i^1, t_i^0, t_i^2\right\}$ 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 sequentially rational strategy for t_j^i is its best-response 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 $\eta_i \in \Delta(A_j)$ over that type's behavior. The resulting optimization problem for type t_i^0, with conjectures η_i over t_j^0's action, is therefore to choose $a_i' \in A_i$ that maximizes the following expected payoff:

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

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

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

□

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 important 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, $R_i:T_i^*\rightrightarrows S_i$.[6]

Lemma 2 (Upper-hemicontinuity) *$R_i:T_i^*\rightrightarrows S_i$ is an* upper-hemicontinuous *correspondence. That is: for any $t_i\in T_i^*$, any $s_i\in S_i(t_i)$ and any sequence $(t_i^n)_{n\in\mathbb{N}}$ in T_i^*, if $t_i^n\to t_i$ and $s_i\in R_i\left(t_i^n\right)$ for every $n\in\mathbb{N}$, 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

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

space. This is a robustness property in that 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, in which it is customary to identify *robustness* with u.h.c.. For instance, WY's unrefinability results (respectively, Penta's (2012)) can be summarized by saying that ICR (resp., ISR) is the strongest robust solution concept among its refinements. As we will show shortly, however, whereas R_i is robust in this sense with the extensive-form uncertainty we consider here it will not be the *strongest* robust solution concept on T_i^*: a proper refinement of R_i is also robust in our space.

3 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 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-responses. 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.[7]

[7]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

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 (see Example in Section 1.1); second, best-responses must account for the 'hybrid' problems illustrated in Ex.1.[8] 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 $k \geq 1$,

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

where we recall that $a_j^*(\cdot)$ denotes j's sequential best response, as a function of a_i. The requirement that the argmax in the definition of $\mathcal{B}_i^k$ be *equal* to the singleton that contains a_i, formalizes the idea that actions that get added to the $\mathcal{B}$-sets must be a *unique* best-response to some conjecture, thereby mimicking the role they play in the infection argument we discussed earlier.

Since A 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 j observing his choice a_i, and hence respond by choosing $a_j^*(a_i)$. 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 $\mathcal{B}_j^k$.[9] The following example

the richness assumption.

[8]Another difference is that types in our setting also incorporate information. With the exception of Penta (2012), which allowed for information partitions with a product structure, the papers cited in the previous footnote (as well as Weinstein and Yildiz (2007, 2011, 2013, 2016)) maintain that types have no information.

[9]The definition of $\mathcal{B}_i^k$ may seem to incorporate an implicit assumption of independence between player i's beliefs about the observability of his action (p), and his beliefs about j's choice (η_i) in the event in which a_i is not observable. But since player j is already assumed to play $a_j^*(a_i)$ whenever a_i is observable (which happens with probability $(1-p)$), such

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 η_1 attaches probability one to $C \in \mathcal{B}_2^2$, and let $p \in [0,1]$. Then, the expected payoffs from player 1's actions are:

$$EU_1(U; p, \eta_1) = p \cdot 0 + (1-p)\,4 = 4 - 4p$$
$$EU_1(M; p, \eta_1) = p \cdot 2 + (1-p)\,2 = 2$$
$$EU_1(D; p, \eta_1) = p \cdot x + (1-p)\,3 = 3 - (3-x)\,p$$

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 $\mathcal{B}_2 = \{L, C\}$. □

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. Exploiting again Assumption 1, which ensures that

independence assumption entails no loss of generality in this case.

$a_i^*(\cdot)$ is the only sequentially rational strategy for types who move second, for those types it is convenient to initialize the procedure directly from this point. Formally: for each i and t_i, let

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

Then, for all $k \geq 1$, having set $RP_j^{k-1} := \{(s_j, t_j) : s_j \in RP_j^k(t_j)\}$, we have,

$$RP_i^k(t_i) := \left\{ s_i \in RP_i^{k-1}(t_i) : \begin{array}{l} \exists \mu_i \in \Delta(RP_j^{k-1} \times \Omega) \text{ such that:} \\ (i) \quad \mathrm{marg}_{T_j \times \Omega} \mu_i = \tau_i(t_i), \\ (ii) \quad s_i \in \arg\max_{s_i' \in S_i(t_i)} \sum_{\omega \in \Omega} \sum_{s_j \in S_j(\omega)} \mu_i[\{(s_j, \omega)\} \times T_j] \, u_i(s_i', s_j, \omega) \end{array} \right\}.$$

Finally, set $RP_i(t_i) := \bigcap_{k \geq 0} R_i^k(t_i)$. It is immediate to check that the solution concept RP_i coincides with R_i whenever $\mathcal{B} = A$, and hence the 'robust predictions' would be no finer than R_i itself in that case. In general, however, $RP_i(t_i) \subseteq R_i(t_i) \cap \mathcal{B}_i$ for all t_i such that $\hat{\theta}_i(t_i) = \theta_i'$, whereas $RP_i(t_i) = \{a^*(\cdot)\} = R_i(t_i)$ for all t_i such that $\hat{\theta}_i(t_i) = \theta_i''$. Hence, RP_i is formally a refinement of R_i, and a proper one if, for instance, $\mathcal{B}_i \subsetneq R_i(t_i)$ for some t_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 generically coincide and are single-valued:

Theorem 1 (Robust Predictions) *For any player i, $RP_i : T_i^* \rightrightarrows A_i$ is nonempty-valued and upper-hemicontinuous. Moreover, for any finite type $t_i \in T_i^*$, and for any strategy $s_i \in RP_i(t_i)$, there exists a sequence of finite types $(t_i^n)_{n \in \mathbb{N}}$ in T_i^*, with limit t_i, and such that $R_i(t_i^n) = RP_i(t_i^n) = \{s_i\}$ for every $n \in \mathbb{N}$.*

The first part of Theorem 1 ensures that the predictions of RP_i are nonempty and 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 second 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 and RP_i for some finite type arbitrarily close to t_i. This has a few important implications: (i) first, RP_i is the *strongest* robust refinement of R_i, since no refinement of RP_i 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 a strongest u.h.c. refinement of R_i, 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 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 *The following hold:*

(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 and any $s_i \in S_i(t_i)$, if there exists a sequence $(t_i^n)_{n\in\mathbb{N}}$ in T_i^* with limit t_i such that $R_i(t_i^n) = \{s_i\}$ for every $n \in \mathbb{N}$, then $s_i \in RP_i(t_i)$*

Hence, while there is a clear formal similarity between Theorem 1 and the 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. 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 $RP(t^{CB}(\omega^0)) = R(t^{CB}(\omega^0)) \cap \mathcal{B} = \{(M, C)\}$.

We also note an obvious but interesting nonmonotonicity of the set of robust predictions: for instance, if U were dropped from this game, then the rationalizable set under common belief would not be affected, but the Stackelberg profiles would be $a^1 = (D, R)$ and $a^2 = (M, C)$. It follows that $\mathcal{B} = \{M, D\} \times \{C, R\}$ and it is easy to show that $RP\left(t^{CB}\left(\omega^0\right)\right) = \{M, D\} \times \{C, R\}$. Hence, eliminating actions from a game may enlarge the RP set. □

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 nowhere dense sets of types. Then, part (i) of Corollary 1 implies that whereas RCBR may deliver less sharp predictions than RP for nongeneric 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 generically on the universal type space. In this sense, RP_i characterizes the 'regular predictions' of RCBR. Formally:

Definition 1 *For any type* $t_i \in T_i^*$ *and any strategy* $s_i \in R_i(t_i)$, s_i *is a* regular prediction of RCBR *for* t_i *if for any neighborhood* $\mathcal{N}(t_i)$ *of* t_i, *there exists an open set* $\mathcal{U} \subseteq \mathcal{N}(t_i)$ *such that* $s_i \in R_i(t_i')$ *for every* $t_i' \in U$.

Corollary 2 *For any type* $t_i \in T_i^*$ *and any strategy* $s_i \in R_i(t_i)$, s_i *is a* regular prediction of RCBR *for* t_i *if and only if* $s_i \in RP_i(t_i)$.[10]

Hence, the RP solution concept is the answer to our opening question: it characterizes the predictions that an analyst could make, for instance in a 'standard' model (i.e., one which maintains standard common knowledge assumptions on the extensice form), to capture the strategic implications of a situation in which players entertain higher order uncertainty over the observability of their actions. As we will show in the next section, such robust predictions can prove especially insightful in a few important classes of games.

Theorem 1, however, is obviously predicated under the assumption that the underlying payoffs are common knowledge. This is useful to distill the specific implications of higher-order uncertainty about the observability actions. However, one should be cautious in just taking RP as a robust solution concept, *tout court*: presumably, players in reality may face higher order uncertainty about both the observability of actions, and their payoffs – which, in the language of the earlier discussion, would lead to a substantially richer set of seeds.

One may thus wonder how the results in Theorem 1 would be affected if uncertainty over the observability of actions interacted with payoff uncertainty. It can be shown that, as long as the added payoff states satisfy a slight strengthening of Assumption 1, the result of Theorem 1 would still go through, with the only difference that the sets $\mathcal{B}$ may grow larger (though not necessarily), and hence entail weaker robust predictions. For example, if one added a richness condition à la WY, then trivially $\mathcal{B}_i = A_i$, and hence

[10] We note that the open sets $\mathcal{U}$ 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.

the strongest robust predictions around the common-belief 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 makes sense to consider uncertainty over utility functions v_i. In most economic applications, however, it would still be sensible to maintain common knowledge that such v_i are increasing. But note that, even if we took the space of payoff uncertainty to include all possible profiles of such functions, the sets $\mathcal{B}$ (and, hence, the robust predictions) would still not be affected. That is because the Stackelberg profiles in that game are pinned down by the ordinal preferences, and no other actions can be made sequential best responses without violating monotonicity of the v_i functions, or also relaxing common knowledge of the outcome function g. (We note that this observation applies to any game which satisfies the conditions of any of the Propositions in Section 4.)

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. But once again, as long as the added states satisfy a strengthening of Assumption 1, it can be shown that the main result goes through unchanged, with the only difference that the sets $\mathcal{B}$ may grow larger (though not necessarily, as we discussed).

4 Applications

In Example 3, not only are the robust predictions particularly sharp, but they also imply that, for a generic set of types, equilibrium coordination arises as the *only* behavior consistent RCBR, i.e., without imposing correctness of beliefs. In Section 4.1 we consider classes of games in which the robust predictions take this especially strong form, and hence equilibrium coordination arises purely from individual reasoning. Section 4.2 explores other classes of games, in which Theorem 1 also has strong implications, which may or may not lead to eductive coordination. Section 4.3 contains our results on environments with one-sided uncertainty, and conditions under which a first-mover advantage is 'pervasive'.

4.1 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, two main reactions can be found in the literature. At one extreme, nonequilibrium approaches such as rationalizability (e.g., Bernheim (1984) and Pearce (1984)) or level-k theories (e.g., Nagel (1995)) have been developed to analyze initial responses in games. At the opposite extreme, other approaches have developed Schelling's (1960) idea of *focal points* (e.g., Sugden (1995)), which maintains the equilibrium assumption and shifts the discussion on the mechanisms that bring about coordination to external properties of the game, which are not included in the extensive form or related to players' payoffs in the game.

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:

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^i_j 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)).*[11]

[11] Formally, a *coordination game* is a game in which every profile in which players choose

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.[12] 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 from Luce and Raiffa (1957) to the following example:[13]

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

---

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 $a \in A$

[12] 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 simultaneously in the first period, and they alternate after that.

[13] 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.

	L		C		R	
U	100	100	$-\varepsilon$	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 $[C]$. Certainly (ii) is not a very convincing argument if player 1 has any reason to think that player 2 will not choose $[C]$. Also, (i) implies a very pessimistic point of view; to be sure, M yields at least [0], but it also yields at most $[\varepsilon]$." (ibid., p. 62). If 1 had any uncertainty that 2 might be playing L in this game, it would be unreasonable to assume he would not play U for sufficiently small 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)). 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 $t_i^{CB}\left(\omega^0\right)$, yielding the eductive coordination result. As shown by Example 3, however, eductive coordination is 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.

4.2 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:[14]

Proposition 2 *If G^* satisfies Assumption 1, both players are indifferent over nonequilibrium profiles, and they strictly prefer any Nash equilibrium to any nonequilibrium 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 $\hat{\theta}_i(t_i) = \theta_i''$.*

Besides including as special cases important classes of games, such as Harsanyi (1981) and Kalai and Samet's (1984) unanimity games, the conditions in Proposition 2 describe broader, interesting class of strategic situations, in which players agree that any Nash equilibrium outcome is better

[14] Formally, a *unanimity* game (cf. Harsanyi (1981) and Kalai and Samet (1984)) is a coordination game (cf. Footnote 11) such that, $\forall i$, $u_i^*(a') = u_i^*(a'')$ for all nonequilibrium profiles a', a''.

than receiving the 'disagreement payoff' associated to any nonequilibrium outcome. In such a class of 'agreement games', the robust predictions only contemplate that players choose one of the actions associated to the Stackelberg profiles, a_i^i or a_i^j. Note that, beyond finiteness, there is no restriction on the number of actions in the baseline game, or on the rationalizable set, which could be arbitrarily large. That the robust predictions involve at most two actions is thus a remarkably sharp refinement for these games.

Proposition 2 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, although 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.*

4.3 One-sided Uncertainty and Pervasiveness 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 $\Omega^{\dagger} := \{\omega^0, \omega^1\}$ (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 $k \geq 1$:

$$\mathcal{B}_1^{\dagger,k+1} := \mathcal{B}_1^{\dagger,k} \cup \left\{ a_1 \in A_1 : \begin{array}{c} \exists p \in [0,1],\, \exists \eta_1 \in \Delta(\mathcal{B}_2^{\dagger,k}) \text{ such that:} \\ \arg\max_{a_1' \in A_1} \left(p \sum_{a_2 \in A_2} \eta_1[a_2]\, u_1^*(a_1', a_2) + (1-p)\, u_1^*(a_1', a_2^*(a_1')) \right) = \{a_1\} \end{array} \right\},$$

$$\mathcal{B}_2^{\dagger,k+1} := \mathcal{B}_2^{\dagger,k} \cup \left\{ a_2 \in A_2 : \exists \eta_2 \in \Delta(\mathcal{B}_1^{\dagger,k}) \text{ s.t. } \arg\max_{a_2' \in A_2} \sum_{a_1 \in A_1} \eta_2[a_1] \cdot u_2^*(a_2', a_1) = \{a_2\} \right\}$$

Note that $\mathcal{B}_i^{\dagger}$ is basically the same as the set $\mathcal{B}_i$ defined in Section 3, except that only a_1^1 is taken as a 'seed', not a_2^2. For each i, we define the 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, $RP_i^{\dagger} : T_i^{\dagger} \rightrightarrows S_i$ is nonempty-valued and upper-hemicontinuous. Moreover, for any finite type $t_i \in T_i^{\dagger}$ and any strategy $s_i \in RP_i^{\dagger}(t_i)$, there exists a sequence of finite types $(t_i^n)_{n \in \mathbb{N}}$ in $T_i^{\dagger}$ with limit t_i and such that $R_i(t_i^n) = RP_i^{\dagger}(t_i^n) = \{s_i\}$ for every $n \in \mathbb{N}$.*

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

Corollary 4 *The following hold:*

(*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 $t_i \in T_i^\dagger$ and any $s_i \in S_i(t_i)$, if there exists a sequence $(t_i^n)_{n\in\mathbb{N}}$ in $T_i^\dagger$ with limit t_i such that $R_i(t_i^n) = \{s_i\}$ for every $n \in \mathbb{N}$, then $s_i \in RP_i^\dagger(t_i)$*

(*iv*) *For any $t_i \in T_i^\dagger$ and any $s_i \in S_i(t_i)$, s_i is a* regular prediction of RCBR *for type t_i if and only if $s_i \in RP_i^\dagger(t_i)$.*

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) independently 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. 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.[15] 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

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

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 $t^{CB}\left(\omega^1\right)$, and hence locally robust in our model.

From a broader perspective, this result has important implications in relation with the idea, which has received strong support by the experimental literature, that timing and commitment may have strategic importance beyond actual observability of actions. Cooper et al (1993), for instance, have shown 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 also Camerer (2003), and references therein).[16] As we discussed in the introduction, this is in line with the *Kreps hypothesis* (Kreps (1990)), 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.[17] This is not to say that the logic of our results necessarily provides a behaviorally accurate model of individuals' strategic reasoning (see, e.g., Crawford et al. (2013) and references therein), but only that, once combined with this kind of uncertainty, standard assumptions such as RCBR may provide an effective *as if* model of how timing impacts individuals' strategic behavior.

Finally, note that the result in Proposition 3 implies that, with one-sided

[16]In Cooper et al.'s (1993) experimental results, 62% of the row players and 65% of the column players choose the actions associated to their favorite equilibrium in the simultaneous moves version of the Battle of the Sexes, whereas in the sequential version in which row players choose first, followed by column (who still do not observe row's choice), the figures change to 88% and 30%, respectively.

[17]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.

uncertainty, higher-order uncertainty over the observability of actions yields eductive coordination even in games which do not satisfy the condition of Proposition 1.

5 Related Literature

On Perturbations of Common Knowledge: Several papers have studied perturbations of common knowledge assumptions on payoffs, following the seminal paper by Lipman (2003) and Weinstein and Yildiz (2007, WY). WY, in particular, characterize the correspondence of Dekel et al.'s (2007) interim correlated rationalizability (ICR) on the univeral type space generated by a space of payoff uncertainty which satisfies a richness condition for static games (namely, for each player's action, it contains a payoff state at which that action is strictly dominant). They show that ICR is generically single-valued in this space, and whenever it admits multiple rationalizable outcomes for some belief hierarchy, any of those outcomes is uniquely rationalizable for an arbitrarily close sequence of types.

The key insights of WY have been applied to mechanism design by Oury and Tercieux (2012) and the analysis has been extended to dynamic games by Weinstein and Yildiz (2011, 2016), Chen (2012) and Penta (2012). The latter paper also allows for information types, and characterizes the strongest robust predictions in general information partitions with a product structure, under an extensive form richness condition. Penta (2013) relaxes the richness condition in static games, and studies sufficient conditions for Weinstein and Yildiz's selection *without* richness; Chen, Takahashi and Xiong (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.

All these papers, as well as ours, exploit infection arguments similar to Rubinstein's (1989) email game, and which are also common in the contagion literature (e.g., Morris (2000), Steiner and Stewart (2008)), and in global games (e.g., Carlsson and Van Damme (1993), Morris and Shin (1998), Frankel, Morris and Pauzner (2003), Mathevet and Steiner (2013), etc.). As discussed, these arguments typically consist of two main ingredients: the 'seeds' of the infection, and a 'chain of strict best replies'. Our argument differs from the earlier work both in that it relies o less 'seeds' (only the stackelberg actions in our case; all players' actions for the papers based on a richness assumption), and in the chain of best replies (a hybrid of static and sequential best replies in our case; either one or the other in the earlier papers). These differences yield to a structure theorem which, while displaying important similarities with WY's, at the same time describes a very different correspondence: In both cases, multiplicity is only possible within nowhere dense 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 nowhere dense sets of belief hierarchies. No analogs of this phenomenon can be found in WY's space.

It can be shown that our exercise can be mapped to one of payoff uncertainty for a properly designed artificial *auxiliary game*. 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 (Penta and Zuazo-Garin (2019)). Thus, none of the existing results can be directly applied to the auxiliary game. Yet, it may still be tempting to think that the existence of an u.h.c. refinement of R should perhaps be expected (lack of richness afterall entails a smaller set of perturbations than in WY, thereby making it easier to preserve continuity).[18] But the fact that both R and RP generically coincide and are single-valued is *not* a direct implication of the lack of richness: without richness, payoff perturbations alone would

[18]Penta (2013), however, cautioned against this perhaps natural conjecture, by showing that weak conditions on a space of payoff uncertainty without richness may entail exactly the same structure theorem as WY.

often induce open sets of types with multiple rationalizable actions.

On Extensive-Form Uncertainty: A few papers have studied models with uncertainty over the observability of actions. Robson (1994), in particular, introduced a refinement for two-player non zero-sum games, using the same set of states and information partition as in our model. On a similar vein, Reny and Robson (2004) model a situation in which players' 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, however, 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. Solan and Yariv (2004) studied a game in which the monitoring structure is endogenous, and commonly known in equilibrium. Zuazo-Garin (2017) introduced 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.

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

6 Conclusions

In this paper we studied the implications of perturbing common knowledge assumptions on the observability of actions in two-player games. 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*) efficient coordination in common interest games; (*iv*) 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 of those in the previous paragraph, 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, 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 problem of extensive-form uncertainty is very broad, and little understood. In this paper we have focused on one particular form it can take, but of course more research is needed to address the broader question. In Section 3 we discussed how our results extend to environments with payoff uncertainty, as well as to richer extensive form uncertainties. An important and more challenging extension would be to games with more than two players, which would require dealing with the richness of extensive forms associated to a larger set of players. 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., 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 uncertainty, and Bergemann and Morris (2005, 2009) and Penta (2015) in mechanism design). 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 broad and complex. We provided one of the first attempts at a systematic understanding of this question, and we have shown that basic and plausible forms of extensive-form uncertainty may deliver novel qualitative insights on important classes of games, which include many archetypal models of both conflict and coordination, as well as on many classical questions (such as the Kreps Hypotehsis, the tension between maxmin and EU-maximization in zero-sum games, efficient coordination in common interest games). But the modeling possibilities are very rich, and our results as well as the richness of such possibilities suggest that further exploring the problem of extensive-form uncertainty may prove to be a fertile direction for future research.

Appendix

A The Universal Type Space

In this appendix we recall the construction of the universal type space informally introduced in Section 2, obtained minimally adapting the standard one by Brandenburger and Dekel (1993). The construction for the space in Section 4.3 proceeds in analogous way, applying the obvious changes in set of states Ω and the information types in each Θ_i. Conceptually, the elements of the universal type space formalizes players' belief-hierarchies in a specific way. That is, for every i, his beliefs about Ω (*first order beliefs*), his beliefs about Ω and the opponent's first order beliefs (*second order beliefs*), and so on. Remember that each $\theta_i \in \Theta_i$ is a subset of Ω. Then, the set of possible *first order beliefs consistent with* θ_i is defined as $Z_i^1(\theta_i) := \Delta(\theta_i)$. Also define player j's first order beliefs that are consistent with i's information θ_i as:

$$Z_j^1\left(\theta_i\right) := \left\{\pi_j^1 \in Z_j^1(\theta_j) : \theta_j \cap \theta_i \neq \emptyset\right\}.$$

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

$$Z_i^{k+1}(\theta_i) := \left\{\left(\pi_i^\ell\right)_{\ell=1}^{k+1} \in Z_i^k\left(\theta_i\right) \times \Delta\left(\theta_i \times Z_j^k\left(\theta_i\right)\right) : \mathrm{marg}_{\theta_i \times Z_j^{k-1}(\theta_i)} \pi_i^{k+1} = \pi_i^k\right\},$$

$$Z_j^{k+1}(\theta_i) := \left\{\left(\pi_j^\ell\right)_{\ell=1}^{k+1} \in Z_j^{k+1}(\theta_j) : \theta_j \cap \theta_i \neq \emptyset\right\}.$$

The *first order beliefs* of a type with information θ_i are elements of $\Delta(\theta_i)$. An element of $\Delta(\theta_i \times Z_j^{k-1}(\theta_i))$ is the *k-th order belief* of a type with information θ_i. The set of (*collectively coherent*) *belief hierarchies* for type

θ_i is then defined as:

$$H_i(\theta_i) := \left\{ \pi_i \in Z_i^1(\theta_i) \times \prod_{k\in\mathbb{N}} \Delta\left(\theta_i \times Z_j^k(\theta_i)\right) : \forall k \in \mathbb{N}, \left(\pi_i^\ell\right)_{\ell=1}^k \in Z_i^k(\theta_i) \right\},$$

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

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

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

(*i*) $\pi_i^1[\omega] = \tau_i^*(t_i)[\text{Proj}_\Omega^{-1}(\omega)]$ for any state ω.

(*ii*) $\pi_i^{k+1}[E] = \tau_i^*(t_i)[\text{Proj}^{-1}_{\Omega\times\theta_i\times Z_i^k(\theta_i)}(E)]$ for any measurable $E \subseteq \Omega\times\theta_i\times Z_i^k(\theta_i)$ and any $k \geq 1$.

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

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

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

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

- Higher order beliefs specified by, for each $k \geq 1$, map $\hat{\pi}_i^k : T_i \to \Delta(\Omega \times Z_j^k)$, where for any measurable $E \subseteq \Omega \times Z_j^k$,

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

Then, continuous map $\phi_i : T_i \to T_i^*$ given by $t_i \mapsto (\hat{\theta}_i(t_i), \hat{\pi}_i(t_i))$ where $\hat{\pi}_i(t_i) := (\hat{\pi}_i^k(t_i))_{k\in\mathbb{N}}$ assigns to each type in an information-based type space the induced information-hierarchy pair that corresponds. Mertens and Zamir (1985) showed that for arbitrary non-redundant information-based type space $\mathcal{T} = (T_i, \hat{\theta}_i, \tau_i)_{i\in I}$,[19] set $\phi_i(T_i)$ is a *belief-closed* subset of T_i^*, in the sense that for every type $t_i \in \phi_i(T_i)$ belief $\tau_i^*(t_i)$ assigns full probability to $\phi_j(T_j)$. A type $t_i \in T_i^*$ is *finite* if it belongs to a finite belief-closed subset of T_i^*.

B Proof of Theorems 1 and 2

B.1 Proof of Part (*i*)

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 $E \subseteq S_j \times T_j \times \Omega$. 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 $\nu_i'[E] := \tau_i(t_i')[\text{Proj}_{T_j'\times\Omega}(E)]$. 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 $\Omega^\dagger$. ■

[19]Type space $\mathcal{T}$ is *non-redundant* if for any player i map ϕ_i is injective.

B.2 Proof of Part (*ii*)

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 for types with information θ_i'' come directly from Assumption 1, which implies a unique best-reply. 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 $(t_i^n)_{n\in\mathbb{N}}$ with limit t_i and strategy $s_i \in \bigcap_{n\in\mathbb{N}} RP_i^{k+1}(t_i^n)$. For each $n \in \mathbb{N}$ take conjecture μ_i^n that justifies the inclusion of s_i in $RP_i^{k+1}(t_i^n)$. We know from compactness of $\Delta(S_j \times T_j \times \Omega)$ that there exists some convergent subsequence of $(\mu_i^n)_{n\in\mathbb{N}}$, $(\mu_i^{n_m})_{m\in\mathbb{N}}$, 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 $\mu_i[RP_j^k \times \Omega] \geq \text{limsup}_{m\to\infty}\mu_i^{n_m}[RP_j^k \times \Omega] = 1$. 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 $(t_i^n)_{n\in\mathbb{N}}$ such that $\tau_i(t_i^n) = (1 - 1/n) \cdot \tau_i(t_i) + (1/n) \cdot t_i^i$, where t_i^i is the type consistent with common belief in ω^i. Obviously, $(\tau_i^n)_{n\in\mathbb{N}}$ approaches t_i. Moreover, p^n and η_i^n are defined from t_i^n analogously as p and η_i are for type t_i, and hence (using Assumption 1) for n large enough the 'hybrid' best-response is unique. Hence there exist $\bar{n}$ and a_i such that $s_i \in \bigcap_{n\geq\bar{n}} RP_i^{k+1}(t_i^n)$ and thus $s_i \in RP_i^{k+1}(t_i)$ from u.h.c. of RP_i^{k+1}. ■

B.3 Proof of Part (*iii*)

The proof exploits the following auxiliary solution concept: For each type t_i let $W_i^{\mathcal{B}}(t_i) := \bigcap_{k\geq 0} W_i^{\mathcal{B},k}(t_i)$, where $W_i^{\mathcal{B},0}(t_i) := \mathcal{B}_i$ if t_i has information θ_i' and $W_i^{\mathcal{B},0}(t_i) = \{a_i^*(\cdot)\}$ otherwise, and then, for every $k \geq 0$, having defined $W_j^{\mathcal{B},k} = \{(s_j, t_j) : s_j \in W_j^{\mathcal{B},k}(t_j)\}$, we have:

$$W_i^{\mathcal{B},k+1}(t_i) := \left\{ s_i \in W_i^{\mathcal{B},k}(t_i) : \begin{array}{c} \exists \mu_i \in \Delta(W_j^{\mathcal{B},k} \times \Omega) \text{ such that:} \\ (i) \quad \text{marg}_{T_j\times\Omega}\mu_i = \tau_i(t_i) \\ (ii) \quad \arg\max_{s_i' \in S_i(t_i)} \sum_{\omega\in\Omega} \sum_{s_j \in S_j(\omega)} \mu_i[\{(s_j,\omega)\} \times T_j] \cdot u_i(s_i', s_j, \omega) = \{s_i\} \end{array} \right\}.$$

Lemma 3 *For every $k \geq 0$, every player i, every state ω and every strategy $s_i \in \mathcal{B}_i^k$, if $\omega \in \theta_i'$, and $s_i \in \{a_i^*(\cdot)\}$ otherwise, there exists some finite type $t_i^{s_i,\omega} \in T_i^*$ with information $\theta_i(\omega)$ such that $R_i^{k+1}(t_i^{s_i,\omega}) = \{s_i\}$.*

Proof. We proceed by induction on k. The initial step ($k = 0$) holds trivially: for states $\omega \in \theta_i'$ let $t_i^{a_i^i,\omega}$ be the type that represents common belief in ω^i, and for state ω^j let $t_i^{a_i^*(\cdot),\omega^j}$ be the type that represents common belief in ω^j. For the inductive step, let $k \geq 0$ be such that the claim holds; we verify that it also holds for $k+1$. Fix strategy s_i in the appropriate set and conjecture η_i which justifies the inclusion of s_i on said set. We know from the inductive hypothesis that $\forall (s_j, \omega) \in \text{supp } \eta_i$, there exists some finite type $t_j^{s_j,\omega}$ with information $\theta_j(\omega)$ and for which $R_j^{k+1}(t_j^{s_j,\omega}) = \{s_j\}$. Define $t_i^{s_i,\omega}$ as the type with information θ_i and beliefs $\tau_i[E] := \eta_i[\{(s_j,\omega) \in S_j \times \Omega : (t_j^{s_j,\omega}, \omega) \in E\}]$ for every measurable $E \subseteq T_j$. Obviously, $t_i^{s_i,\omega}$ is well-defined and finite. Pick now arbitrary conjecture μ_i that puts probability one on the graph of R_j^{k+1} and is consistent with $t_i^{s_i,\omega}$. Notice that for every $(s_j, \omega') \in \text{supp } \eta_i$ we have that:

$$\mu_i\left[T_j \times \{(s_j,\omega')\}\right] = \mu_i\left[\left\{t_j^{s_j,\omega''} : \omega' \in \theta_j(\omega'') \cap \theta_i(\omega)\right\} \times \{(s_j,\omega')\}\right]$$

$$= \mu_i \left[\left\{ t_j^{s_j, \omega''} : \omega' \in \theta_j(\omega'') \cap \theta_i(\omega) \right\} \times S_j \times \{\omega'\} \right]$$
$$= \tau_i \left[\left\{ t_j^{s_j, \omega''} : \omega' \in \theta_j(\omega'') \cap \theta_i(\omega) \right\} \times \{\omega'\} \right] = \eta_i[(s_j, \omega')].$$

Clearly, it follows that $R_i^{k+2}(t_i^{s_i,\omega}) = \{s_i\}$. ■

Lemma 4 *For every i, every finite type $t_i \in T_i^*$ and every $s_i \in RP_i(t_i)$ there exists a sequence of finite types $(t_i^n)_{n\in\mathbb{N}}$ in T_i^* converging to t_i such that $s_i \in W_i^{\mathcal{B}}(t_i^n)$ for every $n \in \mathbb{N}$.*

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

- Set of types $T_i^n := \{n\} \times \{(s_i, t_i), (s_i, t_i^{s_i}) : t_i \in T_i \text{ and } s_i \in RP_i(t_i)\}$, where $t_i^{s_i}$ is constructed as in Lemma 3. Obviously, T_i^n is a finite set.
- Information-map $\hat{\theta}_i^n : T_i^n \to \Theta_i$ given by $(n, s_i, t_i) \mapsto \hat{\theta}_i(t_i)$.
- Finally, in order to define belief-maps, for state $\omega \in \theta_i'$ and strategy $s_i \in \mathcal{B}_i$ let $\eta_i^{s_i,\omega}$ denote a conjecture over $S_j \times \Omega$ that justifies the inclusion of s_i in $\mathcal{B}_i$. For state $\omega \in \theta_i''$ and strategy $s_i = a_i(\,\cdot\,)$ let $\mu_i^{s_i,\omega}$ be an arbitrary conjecture over $S_j \times \Omega$ consistent with θ_i''. Then, define player i's belief-map $\tau_i^n : T_i^n \to \Delta(T_j^n \times \Omega)$ as follows:

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

  for every $(n, s_j, t_j, \omega') \in T_j^n \times \Omega$ such that (t_j, ω') is in the support of $\eta_i^{s_i,\omega}$, and $t_j^{s_j,\omega'}$ is constructed as in Lemma 3. The finiteness of the set of types guarantees that these belief-maps are well-defined and continuous, and that every type in T_i^n and T_j^n is finite.

We claim that the following hold: (i) $\forall t_i \in T_i$, each $(n, s_i, t_i)_{n\in\mathbb{N}}$ converges to t_i; (ii) $\forall t_i \in T_i$ and $\forall s_i \in RP_i(t_i)$, $s_i \in W_i^{\mathcal{B}}(n, s_i, t_i)$ for every $n \in \mathbb{N}$. To prove the claim of the lemma, fix player i and pick arbitrary finite type $\bar{t}_i \in T_i^*$ and action $\bar{s} \in RP_i(\bar{t}_i)$. 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 $((T_i^n, \hat{\theta}_i^n, \tau_i^n)_{i=1,2})_{n\in\mathbb{N}}$ constructed above. By type space invariance, $s_i \in RP_i(\hat{t}_i)$ and by the construction above we know that $\forall n \in \mathbb{N}$ there exists some type $t_i^n \in T_i^n$ such that $\bar{s}_i \in W_i^{\mathcal{B}}(t_i^n)$. Let $(\bar{t}_i^n)_{n\in\mathbb{N}}$ the sequence in the universal type space induced by $(t_i^n)_{n\in\mathbb{N}}$. Again, because of type space invariance we know that $\bar{s}_i \in W_i^{\mathcal{B}}(\bar{t}_i^n)$ for every $n \in \mathbb{N}$. Finally, since we know that $(t_i^n)_{n\in\mathbb{N}}$ converges to $\hat{t}_i$ we also know that $(\bar{t}_i^n)_{n\in\mathbb{N}}$ converges to $\bar{t}_i$ and hence, the proof is complete. ■

For the following lemma let $m \in \mathbb{N}$ be such that $\mathcal{B}_i = \mathcal{B}_i^m$ for every player i. Then, we have that:

Lemma 5 *For every $k \geq 1$, every player i, every finite type $t_i \in T_i^*$ and every strategy $s_i \in W_i^{\mathcal{B},k}(t_i)$ 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 $R_i^{m+k+2}(t_i^k) = \{s_i\}$.*

Proof. We proceed by induction on k. For the initial step ($k = 1$) set $\ell = 1$. Fix player i, finite type $\bar{t}_i$, action $\bar{s}_i \in W_i^{\mathcal{B},\ell}(\bar{t}_i)$ and conjecture $\bar{\mu}_i$ that justifies the inclusion of $\bar{s}_i$ in $W_i^{\mathcal{B},\ell}(\bar{t}_i)$. Then, we know by Lemma 3 that $\forall (s_j, t_j) \in \text{supp}(\text{marg}_{S_j\times T_j}\bar{\mu}_i)$, there exists a finite type $t_j^{\ell-1}(s_j, t_j)$ with the same information as t_j and s.t. $R_j^{m+\ell}(t_j^{\ell-1}(s_j, t_j)) = \{s_j\}$. Then, let type t_i^ℓ have information $\hat{\theta}_i(\bar{t}_i)$ and beliefs $\tau_i^\ell[E] := \bar{\mu}_i[\{(s_j, t_j, \omega) \in S_j \times T_j \times \Omega : (t_j^{\ell-1}(s_j, t_j), \omega) \in E\}]$, 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 $R_j^{m+\ell}$ and notice that for every (s_j, ω) we have that:

$$\mu_i[T_j \times \{(s_j, \omega)\}] =$$

$$= \mu_i \left[\left\{ t_j^{\ell-1}(s'_j, t'_j)) : (s'_j, t'_j) \in S_j \times T_j, s_j \in R_j^{m+\ell}(t_j^{\ell-1}(s'_j, t'_j)) \right\} \times \{(s_j, \omega)\} \right]$$

$$=\mu_i \left[S_j \times \left\{t_j^{\ell-1}(s_j', t_j') : (s_j', t_j') \in S_j \times T_j, R_j^{m+\ell}(t_j^{\ell-1}(s_j', t_j')) = \{s_j\}\right\} \times \{\omega\}\right]$$
$$=\tau_i^{\ell} \left[\left\{t_j^{\ell-1}(s_j', t_j') : (s_j', t_j') \in S_j \times T_j, R_j^{m+\ell}(t_j^{\ell-1}(s_j', t_j')) = \{s_j\}\right\} \times \{\omega\}\right]$$
$$=\bar{\mu}_i \left[\left\{(s_j', t_j') \in S_j \times T_j : R_j^{m+\ell}(t_j^{\ell-1}(s_j', t_j')) = \{s_j\}\right\} \times \{\omega\}\right]$$
$$=\bar{\mu}_i[T_j \times \{(s_j, \omega)\}].$$

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

For the inductive 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 3, but due to the induction hypothesis, instead. ■

Proof of Part *(iii)*. Fix finite type $t_i \in T_i^*$ and strategy $s_i \in RP_i(t_i)$. Then, we know from Lemma 4 that there exists a sequence of finite types $(\hat{t}_i^n)_{n \in \mathbb{N}}$ in T_i^* converging to t_i such that $s_i \in W_i^{\mathcal{B}}(\hat{t}_i^n)$ for every $n \in \mathbb{N}$. It follows from Lemma 5 that $\forall n \in \mathbb{N}$ there exists a sequence of finite types $(t_i^{n,k})_{k \in \mathbb{N}}$ in T_i^* converging to $\hat{t}_i^n$ such that $R_i(t_i^{n,k}) = \{s_i\}$ for all $k \in \mathbb{N}$. Thus, if for each $n \in \mathbb{N}$ we set $t_i^n = t_i^{n,k}$, $(t_i^n)_{n \in \mathbb{N}}$ is a sequence of finite types in T_i^* converging to t_i such that $R_i(t_i^n) = \{s_i\}$ for every $n \in \mathbb{N}$. For Theorem 2, simply repeat the argument substituting T_i^*, $\mathcal{B}$ and R_i with $T_i^{\dagger}$, $\mathcal{B}^{\dagger}$ and $R_i^{\dagger}$, respectively. ■

C Applications

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

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

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

Similarly as above, only a_i^i and a_i^j can be a unique best-response 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 $\mathcal{B}^\dagger$ and RP with $RP^\dagger$. ■

References

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

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

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

4. 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.

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

6. Bergemann, D. and S. Morris (2005) “Robust Mechanism Design,” *Econometrica*, 1521–34.

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

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

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

10. Bernheim, D. (1984) “Rationalizable Strategic Behavior,” *Econometrica* 52, 1007–1028.

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

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

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

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

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

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

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

18. 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.

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

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

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

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

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

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

25. 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.

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

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

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

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

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

31. Lipman, B. (2003) "Finite Order Implications of Common Priors," *Econometrica*, 71, 1255-1267

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

47. Penta, A. and P. Zuazo-Garin (2019) "Fear of Leaks and Payoff Uncertainty," *mimeo*.

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

49. 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.

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

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

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

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

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

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

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

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

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

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

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

Пента, А., Суасо-Гарин, П.

Рационализируемость, наблюдаемость и общие знания [Электронный ресурс] : препринт WP9/2021/05 / А. Пента, П. Суасо-Гарин ; Нац. исслед. ун-т «Высшая школа экономики». – Электрон. текст. дан. (480 Кб). – М. : Изд. дом Высшей школы экономики, 2021. – (Серия WP9 «Исследования по экономике и финансам»). – 55 с. (На англ. яз.)

Мы изучаем стратегическое влияние неопределенности игроков высшего порядка на наблюдаемость действий в обычных играх для двух игроков. Более конкретно, мы рассматриваем пространство всех иерархий убеждений, порожденное неуверенностью в том, будет ли игра вестись как статическая игра или с точной информацией. В этом пространстве мы характеризуем соответствие концепции решения, которая отражает поведенческие последствия рациональности и общего убеждения в рациональности (РОУР), где «рациональность» понимается как *последовательная*, когда игрок ходит вторым. Мы показываем, что такое соответствие в общем случае однозначно и что его структура поддерживает надежное уточнение рационализируемости, что часто имеет очень четкие последствия. Например: (i) в классе игр, который включает как игры с нулевой суммой с чистым равновесием, так и координационные игры с уникальным эффективным равновесием, РОУР в целом обеспечивает эффективные результаты равновесия (*развивающая координация*); (ii) в классе игр, который также включает другие хорошо известные семейства координационных игр, РОУР обычно выбирает компоненты профилей Штакельберга (*выбор Штакельберга*); (iii) если общеизвестно, что действие игрока 2 не наблюдается (например, потому что обычно известно, что игрок 1 делает ход раньше, и т.д.) в классе игр, который включает все вышеперечисленные, РОУР обычно выбирает равновесие статической игры, наиболее благоприятной для игрока 1 (широкое распространение преимущества первопроходца).

Пента Антонио, Каталонский фонд исследований и инноваций, Университет Помпеу Фабра и Высшая школа экономики в Барселоне, Испания; E-mail: antonio.penta@upf.edu

Пейо Суасо-Гарин, МИЭФ, Национальный исследовательский университет «Высшая школа экономики», Российская Федерация; E-mail: p.zuazogarin@hse.ru

Препринты Национального исследовательского университета «Высшая школа экономики» размещаются по адресу: http://www.hse.ru/org/hse/wp

Препринт WP9/2021/05
Серия WP9
Исследования по экономике и финансам

Пента Антонио, Пейо Суасо-Гарин

Рационализируемость, наблюдаемость и общие знания

(*на английском языке*)

Изд. № 2339