Robust Multiplicity with a Grain of Naiveté*

Aviad Heifetz† Willemien Kets‡

October 29, 2016

Abstract

Rationalizability is a central concept in game theory. Since there may be many rationalizable strategies, applications commonly use refinements to obtain sharp predictions. In an important paper, Weinstein and Yildiz (2007) show that none of these refinements is robust to perturbations of high-order beliefs. We show that robust refinements do exist if we relax the assumption that all players are unlimited in their reasoning ability. In particular, for a class of models, every strict Bayesian-Nash equilibrium is robust. A researcher interested in making sharp predictions can thus use intuitive principles or models of players' reasoning to select among the strict equilibria of the game, and these predictions will be robust.

*This paper supersedes Heifetz and Kets (2011). We are grateful to the Editor, George Mailath, and two anonymous referees for excellent comments and suggestions that have significantly improved the paper. We thank Sandeep Baliga, Eddie Dekel, Eduardo Faingold, Amanda Friedenberg, Alessandro Pavan, David Pearce, Antonio Penta, Andrés Perea, Marciano Siniscalchi, and Satoru Takahashi for valuable input, and we are grateful to seminar audiences at Hebrew University, Northwestern University, Tel Aviv University, NYU, Penn State, Queen's University, and conference participants at the SING8 conference (Budapest) and the EEA-ESEM conference (Toulouse) for helpful comments. We thank Luciano Pomatto for excellent research assistance.

†Department of Management and Economics, Open University of Israel. E-mail: aviadhe@openu.ac.il. Phone: +972-9-778-1878.

‡Kellogg School of Management, Northwestern University. E-mail: w-kets@kellogg.northwestern.edu. Phone: +1-505-204 8012.

1 Introduction

Rationalizability is a fundamental concept in game theory. As it often yields a large set of predictions, it is common for applications to use refinements. Since modeling a strategic situation inherently involves making strong simplifying assumptions that are satisfied only approximately in reality, it is important that any refinement be robust to slight perturbations of the modeling assumptions. In an important paper, Weinstein and Yildiz (2007) show a surprising negative result: if a researcher cannot observe players' actual higher-order beliefs about payoffs (without any error) and there are no restrictions on payoffs, then refinements cannot eliminate *any* rationalizable strategy. This suggests that if we have only partial knowledge of players' payoff uncertainty, "accounting for incomplete information... casts doubt on all refinements" (Weinstein and Yildiz, 2007, p. 367).

This paper challenges this negative conclusion. We show that refinements can be robust if uncertainty about players' reasoning ability is taken into account. Allowing for uncertainty about players' reasoning ability is natural. Experiments suggest that the standard assumption in game theory that players have an infinite depth of reasoning–i.e., that they form beliefs about payoffs, about others' beliefs about payoffs, about the others' beliefs about their opponents' beliefs, and so on, *ad infinitum*–is an idealization at best (Crawford, Costa-Gomes, and Iriberri, 2013). In many cases, players have a *finite depth of reasoning*, think that others have a finite depth, or think that others think that their opponent has a finite depth, and so on. Assuming that all types have an infinite depth of reasoning, as standard models do, thus constitutes a strong restriction on beliefs. Accordingly, to test the robustness of predictions, a researcher should consider not only perturbations of beliefs about payoffs, but also about reasoning ability.

As we show, standard models assume that players have an infinite depth of reasoning and this is common belief: all players have an infinite depth, believe that others have an infinite depth, etc. Under this assumption (and a richness assumption on the set of possible payoffs), Weinstein and Yildiz (2007) show that if a type has multiple rationalizable actions, then each of these actions can be made uniquely rationalizable by perturbing the type's belief appropriately. This unique prediction is then robust to further belief perturbations. An important implication is that there are no robust refinements of rationalizability in their setting: if we cannot measure the player's type with infinite precision, then for any of the rationalizable actions, we cannot rule out that this action is uniquely rationalizable for the player. So, if a refinement of rationalizability is robust to alternative specifications of beliefs, then it must select each of the rationalizable actions for the type, and the resulting predictions of the refinement are no stronger than those of rationalizability.

This means that, under the assumption that there is common belief in an infinite depth,

there is no scope for refinements if a researcher is concerned with the robustness of his predictions. First, if a type has multiple rationalizable actions, then he cannot robustly select a subset of rationalizable actions. Second, if a type has a unique rationalizable action, then his prediction is robust, but his prediction is determined entirely by mutual beliefs about payoffs. This leaves no room for the researcher to use axiomatic principles (such as payoff dominance) or other criteria (such as those derived from learning or evolutionary models) to further refine his predictions.

This paper challenges both these conclusions. We show that if we slightly depart from standard assumptions and consider environments where players have an infinite depth of reasoning and *almost*-common belief in an infinite depth, then multiplicity is robust: there are types with multiple rationalizable actions such that nearby types have the same rationalizable actions (i.e., the rationalizability correspondence is locally constant at these types). This implies that, unlike in the standard case, mutual beliefs about payoffs do not tie down the predictions of a researcher who is concerned with the robustness of his predictions. On the set of types with robust multiplicity, the researcher can select one of the rationalizable actions, and the resulting prediction is robust. This motivates us to study the robustness of one of the most common refinements of rationalizability, (Bayesian-Nash) equilibrium. We show that for a class of environments, every strict equilibrium is robust.[1]

So, as in the standard case with common belief in an infinite depth, a researcher who is concerned with the robustness of his predictions is limited in his ability to make predictions. However, the challenges in both cases are very different. In the standard case, the only predictions of a refinement that retain their validity when the researcher has only partial information about the players' beliefs are those predictions that are true for all rationalizable strategies. This implies that the researcher cannot obtain sharper predictions than those provided by rationalizability unless he is willing to give up robustness. By contrast, if we allow for uncertainty about players' reasoning ability, there is no such tradeoff: robust refinements of rationalizability exist. However, the robustness requirement alone does not select a particular equilibrium beyond the requirement that incentives are strict: *every* strict Bayesian-Nash equilibrium is robust. To select a particular (strict) equilibrium, the researcher will have to appeal to refinements. In this case, refinements are not just consistent with robustness; they are in fact necessary to make sharp predictions.

This paper is the first to study the robustness of predictions under a larger class of belief perturbations than commonly considered.[2] Unlike the existing robustness literature, which

[1] Of course, not every game has a strict equilibrium. However, the applications we focus on all have multiple strict equilibria.

[2] In a recent paper, Strzalecki (2014) considers the effect of perturbations of beliefs about reasoning ability

considers only perturbations of beliefs about payoffs, we allow for perturbations of beliefs about both payoffs and reasoning ability. We show that allowing for uncertainty about players' reasoning ability has a significant impact on the continuity properties of the rationalizability correspondence and the robustness of predictions, even if the deviation from standard assumptions is small. This is particularly striking given that the characterization results of Weinstein and Yildiz (2007) have otherwise proven to be extremely robust: they extend to dynamic games (Chen, 2012; Weinstein and Yildiz, 2013), games that do not satisfy the richness assumption on payoffs (Penta, 2013; Chen, Takahashi, and Xiong, 2014b), and to general information structures (Penta, 2012).[3] Our results thus suggest that accounting for uncertainty about reasoning ability can lead to novel insights.

The idea that a "grain" of bounded rationality may affect the behavior of rational players has a long history in game theory. Within this literature, our work is most closely related to Kreps, Milgrom, Roberts, and Wilson (1982) and Milgrom and Roberts (1982) who consider the effect of a small amount of doubt about the opponent's rationality. Unlike the irrational types in the existing literature, our nonstrategic types are not committed to taking a certain action. As we discuss in Section 6, this leads to new considerations and insights.

The next section provides an informal overview of our results and puts them in a broader context. The formal treatment starts in Section 3.

2 Preview of main results

2.1 Framework

Standard type spaces model players with an infinite depth of reasoning. As suggested by Harsanyi (1967) and shown formally by Mertens and Zamir (1985), each standard type unfolds into a belief hierarchy with an infinite depth that specifies a player's first-order belief μ^1 (i.e., a probability distribution on the payoff parameters), his second-order belief μ^2 (i.e., his belief about the other player's first-order belief), and so on, *ad infinitum*.

Relaxing this strong assumption requires making the assumptions on players' depth of

in the context of the electronic-mail game of Rubinstein (1989). However, he does not study the robustness of predictions; see Section 6.

[3] A number of authors have shown that the results of Weinstein and Yildiz (2007) do not necessarily extend to other settings. For example, the results do not extend when the topology is changed (Dekel, Fudenberg, and Morris, 2006; Chen, Di Tillio, Faingold, and Xiong, 2010, 2015), when an alternative robustness concept is applied (Chen, Takahashi, and Xiong, 2014a), or in the presence of endogenous information, signaling, and learning (Angeletos and Werning, 2006; Angeletos et al., 2006, 2007; Yang, 2013). Unlike us, these papers do not consider arbitrarily small deviations from standard assumptions.

reasoning explicit within a space of belief hierarchies with an arbitrary (finite or infinite) depth of reasoning. A belief hierarchy has *finite depth* k if it specifies a player's first-order belief μ^1, his second-order belief μ^2, and so on, up to his kth-order belief μ^k but no further. The space of all belief hierarchies (with finite or infinite depth) defines the *universal type space* for players with an arbitrary depth of reasoning, denoted $\mathcal{T}^*$. As in the universal type space $\mathcal{T}^{MZ}$ of Mertens and Zamir (1985) for standard type spaces, every belief hierarchy in $\mathcal{T}^*$ defines a type.[4] With this model in hand, we can characterize the types from standard type spaces:

> **Proposition 3.2 [Characterization of standard types].** *The types from standard type spaces are precisely those types in the universal type space $\mathcal{T}^*$ that have an infinite depth of reasoning and that have common belief in the event that players have an infinite depth of reasoning.*

So, standard types satisfy strong common-knowledge restrictions on their beliefs about players' reasoning ability: not only are players assumed to have an infinite depth of reasoning, they also believe that other players have an infinite depth of reasoning, believe that others believe that, and so on.

Now that these assumptions are made explicit, we can weaken them by considering types in the universal type space that satisfy slighter weaker assumptions. This requires a notion of closeness of beliefs, formally captured by the topology on the type space. The topology reflects what the researcher can learn about the players' types if his observation of their beliefs is imperfect: if a player's actual type is in an open set O and his observation is sufficiently precise, the researcher would conclude that the player's type is in fact in O, even if he may never learn the player's true type. We have in mind a researcher who can observe only finitely many orders of beliefs. So, if the researcher observes a player's beliefs $\mu^1, \ldots, \mu^m$ up to some finite order m, then he finds possible any type whose beliefs $\nu^1, \ldots, \nu^m$ are close to the observed beliefs. On the other hand, he rules out types whose beliefs are very different from the observed beliefs. In particular, he rules out types with a depth of reasoning strictly less than m.

Since applied researchers sometimes restrict attention to a subset of types, we also want our notion of closeness to be independent of the choice of model. For example, if a researcher considers two types to be close when his model contains only types with depth at most k, then he should also consider them close if his model also includes types of higher depth. In particular, if the researcher considers two types to be close when his model is given by the universal type space $\mathcal{T}^{MZ}$ for standard type spaces, then he will also deem them close if his model is the more general model $\mathcal{T}^*$ and vice versa.

[4] The converse also holds: every type (with a finite or infinite depth) corresponds to a type in $\mathcal{T}^*$; see Appendix B.

Together, these two considerations pin down the topology: we use the product topology on the set H^k of types with a given depth k of reasoning, and then "glue" the spaces H^k, $k \leq \infty$, together using the sum topology.[5]

In this topology, types are close to standard types if they have have an infinite depth of reasoning and have mth-order mutual belief in the event that players have an infinite depth of reasoning for some large but finite m. That is, types are close to a standard type if they believe (i.e., assign probability 1 to the event) that players have an infinite depth, they believe that players believe that players have an infinite depth, and so on up to the statement that includes the word "believe" m times, but no further. In that case, we say that there is *almost-common belief in an infinite depth*.

We start by considering interim correlated rationalizability (Dekel, Fudenberg, and Morris, 2007). An action is *(interim correlated) rationalizable* for a type if it survives the iterated elimination of strictly dominated strategies. We assume that a depth-1 type acts as if his opponent is nonstrategic and can play any action. This is in the spirit of the level-k literature, which assumes that a level-1 type plays a best response against a nonstrategic level-0 type that chooses its action uniformly at random (Crawford, Costa-Gomes, and Iriberri, 2013).

A researcher who cannot measure players' belief hierarchies with infinite precision may want his prediction to be robust against small perturbations. To capture this, say that a subset A'_i of actions is *robustly rationalizable* for a type h_i if, when the player's actual type is h_i, then the researcher would conclude that i's rationalizable actions are the actions in A'_i whenever he can measure i's belief with sufficient (but finite) precision. As the topology reflects what a researcher can learn about the players' types, this is the case precisely if there is a neighborhood of h_i (i.e., an open subset $O(h_i)$ that contains h_i) such that the set of rationalizable actions is A'_i across all types in the neighborhood.

2.2 Robust multiplicity

Since every game-theoretic model is an idealization of the true strategic environment, an important question is whether predictions are robust to relaxing strong assumptions embodied in the model. The case of complete-information games has received particular attention in the literature. Complete-information models are an idealization of situations where payoffs are observed only with some noise. The predictions of complete-information models may not be robust to the introduction of a small amount of incomplete information, as the following example illustrates.

[5] The sum topology preserves the open sets in the component spaces without adding extraneous open sets: it is the weakest (i.e., smallest) topology that contains the open sets in H^k, $k \leq \infty$.

Example 1. Consider the following payoff matrix, taken from Carlsson and van Damme (1993):

	I	NI
I	θ, θ	$\theta - 1, 0$
NI	$0, \theta - 1$	$0, 0$

where $\theta \in [-1, 2]$. Players can choose to invest (i.e., play I) or to not invest (i.e., play NI). Each player i receives a (potentially noisy) signal x_i about the state: if the state is θ, then each player receives a signal x_i drawn uniformly at random from $[\theta - \varepsilon, \theta + \varepsilon]$, independently across players, where $\varepsilon \geq 0$ is small (say, $\varepsilon < \frac{1}{2}$). So, players' observations of θ become increasingly precise as ε approaches 0. If there is complete information about payoffs (i.e., $\varepsilon = 0$), then both actions are rationalizable for any signal $x_i \in (0, 1)$.

This multiplicity may not be robust, however. Assuming that players have an infinite depth of reasoning and this is common belief, Carlsson and van Damme (1993) show that the risk-dominant equilibrium is the unique prediction if the noise is small (i.e., ε positive but close to 0). So, in the standard model, the prediction that both actions are rationalizable is not robust to relaxing the assumption that $\varepsilon = 0$. $\triangleleft$

There is thus a striking discontinuity between the case where the payoffs are common belief (i.e., $\varepsilon = 0$) and where they are almost-common belief (i.e., $\varepsilon > 0$), at least if players have an infinite depth of reasoning and this is common belief. This type of sensitivity is very general:

> **Proposition 3.6 [No robust multiplicity with common belief in an infinite depth; Weinstein and Yildiz, 2007, Prop. 2]** *If the set of possible payoff functions is sufficiently rich, then robust multiplicity is not consistent with common belief in an infinite depth of reasoning. That is, there there are no types with multiple rationalizable actions such that nearby types have the same set of rationalizable actions.*

Weinstein and Yildiz's result holds very generally. Thus, there seems to be no hope to have robust multiplicity in a standard type space unless one is willing to make common-knowledge assumptions on the payoff functions (such as $\varepsilon = 0$). However, by working with standard type spaces, Weinstein and Yildiz do make the strong assumption that players have an infinite depth of reasoning and have common belief in an infinite depth. Our first main result shows that multiplicity is robust when this strong assumption is relaxed.

> **Theorem 4.1. [Robust multiplicity with almost-common belief in an infinite depth].** *If the set of possible payoff functions is sufficiently rich, then*

robust multiplicity is consistent with an infinite depth of reasoning and almost-common belief in an infinite depth. That is, given a set A' of actions with $|A'| > 1$, there exist types h^m, $m = 1, 2, \ldots$, with an infinite depth of reasoning and mth-order mutual belief in an infinite depth for whom A' is robustly rationalizable.

We explain the intuition behind Theorem 4.1 using a variant of Example 1.

Example 2. Players believe that the state θ is either $\theta = 2$ or $\theta = -1$. So, the possible payoff matrices are

	I	NI
I	$2, 2$	$1, 0$
NI	$0, 1$	$0, 0$

$\theta = 2$

	I	NI
I	$-1, -1$	$-2, 0$
NI	$0, -2$	$0, 0$

$\theta = -1$

In this case, investing is a strict best response for any player who assigns probability $p > \frac{2}{3}$ to $\theta = 2$, and not investing is a strict best response for a player if he assigns probability $p < \frac{1}{3}$ to $\theta = 2$. If $p \in (\frac{1}{3}, \frac{2}{3})$, then either action is a strict best response for a player depending on his conjecture about the opponent's behavior: under the conjecture that the opponent invests, investing is the unique best response; and under the conjecture that the opponent does not invest, not investing is the unique best response. The probability distributions that assign probability $p \in (\frac{1}{3}, \frac{2}{3})$ to $\theta = 2$ define what we call a *multiplicity set*.

To show that multiplicity can be robust for a type with almost-common belief in an infinite depth, we start with a "grain" of robust multiplicity and use a contagion argument to show that multiplicity is robust for types with almost-common belief in an infinite depth.

The "grain" consists of finite-depth types. We start with the depth-1 types. Depth-1 types form beliefs only about the payoff parameter θ, and act as if their opponent can play any action. Both actions are rationalizable for a depth-1 type h^1 with a belief in the multiplicity set (i.e., that assign probability $p \in (\frac{1}{3}, \frac{2}{3})$ to $\theta = 2$), and the same is true for depth-1 types with beliefs sufficiently close to h^1. So, both actions are robustly rationalizable for type h^1. Then, by a similar argument, both actions are robustly rationalizable for a depth-2 type h^2 with a belief in the multiplicity set that believes that the opponent's type is h^1 or some nearby type for whom both actions are rationalizable. We can iterate this argument to show that for any $k = 1, 2, \ldots$, there is a depth-k type h^k for whom both actions are robustly rationalizable.

This allows us to show that multiplicity can be robust under almost-common belief in an infinite depth. Consider a type with an infinite depth with a belief in the multiplicity set that believes that the opponent has a finite-depth type for whom both actions are robustly rationalizable. By a similar argument as before, both actions are robustly rationalizable for

the type. Again, by iterating the argument, we can show that for any $m = 1, 2, \ldots$, there are infinite-depth types with mth-order mutual belief in the event that players have an infinite depth for whom both actions are robustly rationalizable. So, multiplicity can be robust under almost-common belief in an infinite depth. $\triangleleft$

The proof has two main ingredients: a "grain" of robust multiplicity and a contagion argument. The grain of robust multiplicity consists of finite-depth types that can rationally play any action. A contagion argument then shows that if there is a set V of types for whom both actions are robustly rationalizable, then both actions are robustly rationalizable for any type with beliefs in the multiplicity set that believes (i.e., assigns probability 1 to the event) that the opponent's type is in V. The argument for the general case (Lemma 4.3) requires some care, but the basic insight is simple: if a type has a belief in the multiplicity set, then either action is a strict best response for the type depending on its conjecture about the other player's behavior; and if the type believes that both actions are rationalizable for the opponent, then it can entertain any conjecture. So, by relaxing the assumption that it is common belief that players have an infinite depth, we can introduce a "grain" of robust multiplicity, i.e., finite-depth types for whom both actions are robustly rationalizable. A contagion-type argument then allows us to show that multiplicity is robust even if players have an infinite depth of reasoning and this is almost-common belief.

This implies that even if for types h that are arbitrarily close to the standard case, the rationalizability correspondence may be *finitely determined* at h in the sense that there is some finite n such that for all types close to h, the set of rationalizable actions is unaffected by perturbations of beliefs at order greater than n. So, if a researcher can observe the nth-order beliefs of players (possibly with some noise), then his prediction will be robust. This is not the case if there is common belief in an infinite depth: as is well-known, the rationalizability correspondence is sensitive to perturbations at arbitrarily high order.

While Theorem 4.1 shows that robust multiplicity is consistent with almost-common belief in an infinite depth in a wide range of situations, it does not speak directly to the discontinuity of behavior in Example 1. This is because it leaves open the possibility that the beliefs of the types with robust multiplicity are not consistent with the information structure in the example. However, by focusing specifically on the types with beliefs consistent with the information structure, we can show that multiplicity is robust also for those types:

Theorem 4.7. [Robust multiplicity around complete-information types]. *For every $m = 1, 2, \ldots$, there is an interval $(\underline{x}^{\varepsilon}_m, \bar{x}^{\varepsilon}_m) \supsetneq \{\frac{1}{2}\}$ such that both actions are robustly rationalizable for every infinite-depth type with signal $x_i \in (\underline{x}^{\varepsilon}_m, \bar{x}^{\varepsilon}_m)$ and mth-order mutual belief in an infinite-depth provided that ε is sufficiently small.*

Moreover, $\underline{x}_m^\varepsilon \to 0$ *and* $\bar{x}_m^\varepsilon \to 1$ *as* $\varepsilon \to 0$.

Theorem 4.7 says that as the noise level ε goes to 0, both actions are rationalizable for any type with almost-common belief in an infinite depth that has a signal $x_i \in (0, 1)$. So, the discontinuity in behavior in Example 1 is not robust to relaxing the assumption that there is common belief in an infinite depth. In particular, the risk-dominant equilibrium selection of Carlsson and van Damme (1993) does not extend if we relax the assumption that there is common belief in an infinite depth. Theorem 4.7 thus complements the existing literature: while the literature has shown that the risk-dominant selection is not robust to perturbations of the information structure, the present result shows that the risk-dominant selection is not robust even if we keep the information structure fixed.[6]

Remark 1. Thus far, we have not specified the richness requirement on the set of possible payoff functions. As we discuss, there are different richness assumptions that may be of interest (Assumptions R-Dom and R-Mult(A') below). The interesting case is when the set of possible payoff functions is rich in both senses. This is the case in all examples as well as in most applications in the literature (Morris and Shin, 2003). ◁

2.3 Robust refinements

An important implication of the lack of robust multiplicity in the standard case is that there is no scope for robust refinements if there is common belief in an infinite depth. For example, suppose the payoff matrix is as in Example 1 and $\theta \in (0, \frac{1}{2})$ is commonly known (i.e., $\varepsilon = 0$). Then, both actions are rationalizable, and a researcher who subscribes to payoff dominance may want to select the equilibrium in which both players invest. But, by the results of Carlsson and van Damme (1993), this prediction is not robust: if we introduce a small amount of uncertainty about payoffs, then the not-invest equilibrium is uniquely selected. By the results of Weinstein and Yildiz (2007), the prediction that both players will not invest is also not robust: if we perturb beliefs a little, then the unique prediction is that both players invest. Weinstein and Yildiz (2007) show that this holds very generally: if there is common belief in an infinite depth, then a prediction of a refinement is robust if and only if it is true for *all* rationalizable strategies. So, in that case, a researcher cannot make a prediction that is stronger than what is implied by rationalizability.

If we relax the strong assumption that there is common belief in an infinite depth, then multiplicity is robust, suggesting that there is some scope for robust refinements. To see

[6]See Strzalecki (2014) for a similar result in the context of the electronic-mail game of Rubinstein (1989). However, Strzalecki does not show that the resulting predictions are robust to further belief perturbations.

this, suppose that $\theta \in (0, \frac{1}{2})$ and that ε is close to 0. Suppose a researcher wants to select the payoff-dominant action whenever it is consistent with rationalizability. Then, he could use a refinement of rationalizability that selects the action 'invest' for a type whenever it is rationalizable, and coincides with rationalizability otherwise. This refinement predicts that every type with signal $x_i \in (0, 1)$ whose observation is sufficiently precise (i.e., ε close to 0) will invest. By Theorem 4.7, this selection is robust: even if the researcher has misspecified the model (e.g., the information structure), he can still be confident that players are willing to invest if his assumptions are satisfied approximately. Alternatively, the researcher may want to select the action 'not invest' for a type whenever it is rationalizable (and coincides with rationalizability otherwise). Again, this is a proper refinement of rationalizability, and it is robust.

This motivates us to ask whether standard refinements of rationalizability can be robust. We focus on the robustness of Bayesian-Nash equilibrium, one of the most common refinements of rationalizability. Our equilibrium definition is standard: a strategy profile is a *(Bayesian-Nash) equilibrium* for a model if each type in the model plays a best response to the opponent's strategy. As before, we assume that a depth-1 type plays as if his opponent is nonstrategic and can choose any action, as in the level-k literature (Crawford, Costa-Gomes, and Iriberri, 2013).

Again, we take the perspective of a researcher who can observe finitely many orders of beliefs with some noise: there is some finite order κ and some $\eta > 0$ such that if a player's actual type is h_i, then the researcher cannot rule out any type whose mth-order beliefs are η-close to those of h_i for $m \leq \kappa$, where our notion of η-closeness is determined by the usual weak topology on the set of mth-order beliefs.[7] This leads to the following robustness notion: a Bayesian-Nash equilibrium σ for a model is robust if for some $\eta > 0$ and $\kappa < \infty$, every model that the researcher cannot rule out on the basis of his observations (given η and κ) has an equilibrium σ' such that types that are close to the original model play the same actions as under σ.

The next result shows that for some models, every strict Bayesian-Nash equilibrium is robust:

Theorem 5.2. [Strict equilibrium robust under almost-common belief in infinite depth] *If the set of possible payoff functions is sufficiently rich, then there exist models consistent with an infinite depth of reasoning and almost-common belief in an infinite depth for which every strict Bayesian-Nash equilibrium is robust.*

[7] Recall that mth-order beliefs are probability distributions; so, a sequence $(\mu^{m,n})_n$ of mth-order beliefs converges to an mth-order belief μ^m in the weak topology if it converges in distribution.

Since we can choose the models in Theorem 5.2 so that they include types with multiple rationalizable actions, as we discuss, an immediate implication of Theorem 5.2 is that robust (and proper) refinements of rationalizability exist if we relax the assumption that there is common belief in an infinite depth:

> **Corollary. [Robust refinements under almost-common belief in infinite depth]** *If the set of possible payoff functions is sufficiently rich, then for every $m = 1, 2, \ldots$, there is a model with mth-order belief in an infinite depth for which there is a robust refinement of rationalizability.*

Theorem 5.2 is consistent with a folk result for complete-information games: in environments where payoffs are commonly known among the players, but the researcher is unsure about the payoffs, every strict Nash equilibrium is robust to small misspecifications of the payoffs. Theorem 5.2 shows that this result extends to games with incomplete information with a suitable form of uncertainty about reasoning ability.

We illustrate the intuition behind Theorem 5.2 using Example 1. Suppose that a researcher thinks that there is complete information about payoffs (i.e., $\varepsilon = 0$) and that $\theta \in (0, \frac{1}{2})$. However, he recognizes his model may be misspecified, so he would like his prediction to be robust. For the depth-1 types, who act as if they play against a nonstrategic type, he can select either action. So, let σ^I be a strategy under which all depth-1 types invest; likewise, let σ^{NI} be a strategy under which all depth-1 types do not invest. For $m > 1$, suppose that for $\ell < m$, all depth-ℓ types invest (do not invest) under σ^I (under σ^{NI}). Then, under σ^I, the unique (strict) best response for a depth-m type is to invest; and under σ^{NI}, the unique (strict) best response for a depth-m is to not invest. We can thus construct two strict Nash equilibria, σ^I and σ^{NI}. The former is the payoff-dominant Nash equilibrium in which both players invest, and the latter is the risk-dominant Nash equilibrium in which neither player invests. Since all incentives are strict, these predictions are robust: also if we perturb beliefs a little, each type has a unique best response under either strategy. In particular, these predictions are robust to the introduction of a small amount of incomplete information.

As in the case of rationalizability, equilibrium predictions are not sensitive to the perturbation of beliefs at arbitrarily high order. The argument is slightly different than for our results on rationalizability: whereas the finite-depth types acted as a grain of robust multiplicity in our results for rationalizability, in the proof of Theorem 5.2 they are used to "anchor" the behavior of types with a higher reasoning ability. This makes that the equilibrium correspondence is again finitely determined, just like the rationalizability correspondence. This "anchoring" of equilibrium behavior by finite-depth types is critical: when there is common belief in an infinite

depth, then equilibrium behavior can be sensitive to perturbations of beliefs at arbitrarily high order, as illustrated by the electronic mail game of Rubinstein (1989).

Both in the standard case with common belief in an infinite depth and in the case where there is almost-common belief in an infinite depth, the researcher is thus limited in his ability to make predictions. However, the difficulties he faces are fundamentally different in the two cases. If there is common belief in an infinite depth, the only robust predictions that a researcher can make are the predictions that are true for all rationalizable actions. In particular, equilibrium cannot refine rationalizability if predictions are required to be robust (Weinstein and Yildiz, 2007; also see Proposition 5.1 below). By contrast, if there is almost-common belief in an infinite depth, then robust refinements of rationalizability do exist. However, the requirement that predictions be robust does not select a particular equilibrium; rather, *every* strict equilibrium is robust (Theorem 5.2). In particular, in complete-information games with multiple strict Nash equilibria, all Nash equilibria are robust to the introduction of a small amount of incomplete information about payoffs.

This yields radically different conclusions regarding the scope for the refinement program. Weinstein and Yildiz (2007) argue that since there are no robust refinements of rationalizability when there is common belief in an infinite depth, there is limited or no scope for a refinement program unless a researcher is willing to impose strong common-knowledge restrictions on beliefs (e.g., pp. 374–375). In contrast, the present results suggest that there need not be a tension between robustness and refinements if the strong assumption that there is common belief in an infinite depth is relaxed. In fact, if there are multiple strict equilibria, then a researcher who wants to make sharp predictions *needs* to appeal to refinements to provide sharp predictions, and these refinements will be robust provided that incentives are strict.

The remainder of this paper is organized as follows. Section 3 introduces the framework. Section 4 presents our results on robust multiplicity in the context of rationalizability, and Section 5 presents our results on robust refinements of rationalizability. Section 6 discusses the related literature. Proofs and additional results can be found in the appendices.

3 Framework

3.1 Preliminaries

We follow the standard conventions for subspaces, products, and (disjoint) unions of topological spaces. A subspace of a topological space is endowed with the relative topology, and the

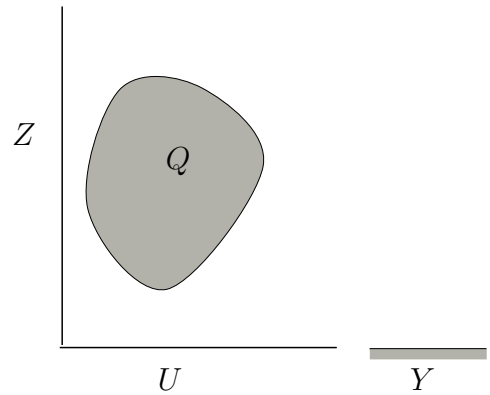

Figure 1: The space W (shaded gray) is the union of $Q \subseteq U \times Z$ and of Y. The space V is the union of U and Y.

product of a collection of topological spaces is endowed with the product topology.[8] If $(V_\lambda)_{\lambda\in\Lambda}$ is a family of disjoint topological spaces, then $\bigcup_\lambda V_\lambda$ is endowed with the sum topology, that is, a subset $U \subseteq \bigcup_{\lambda\in\Lambda} V_\lambda$ is open in $\bigcup_{\lambda\in\Lambda} V_\lambda$ if and only if $U \cap V_\lambda$ is open in V_λ for each $\lambda \in \Lambda$.

Given a topological space V, the set of probability measures on the Borel σ-algebra $\mathscr{B}(V)$ is denoted by $\Delta(V)$. We endow $\Delta(V)$ with the topology of weak convergence. We extend the definition of a marginal to a union of measurable spaces. Let V be the union of the disjoint sets U and Y, and let $Q \subseteq U \times Z$ and $W = Q \cup Y$, where all spaces are assumed to be topological spaces; see Figure 1. Then for $\mu \in \Delta(W)$ denote by $\mathrm{marg}_V \mu \in \Delta(V)$ the probability measure defined by

$$\mathrm{marg}_V \mu(E) = \mu(\{(u, z) \in Q : u \in E\}) + \mu(E \cap Y)$$

for every measurable set $E \subseteq V$. This definition reduces to the standard one if Y is empty. If μ is a probability measure on a product space $U \times Y$, and E is a measurable subset of U, then we sometimes write $\mu(E)$ for $\mathrm{marg}_U\, \mu(E)$.

3.2 Strategic environment

There are two players, labeled $i = 1, 2$. The set of states of nature is Θ. Each player i has a set A_i of actions and a utility function $u_i : A \times \Theta \to \mathbb{R}$. Players may have private information: each player i has a (payoff-irrelevant) signal $x_i \in X_i$. We assume that Θ and X_i are compact metric. Action sets are assumed to be finite, and payoff functions are taken to be continuous. The extension to an arbitrary (finite) number of players is straightforward.

We focus on the case where the set of possible payoff functions is sufficiently rich. One such richness requirement, due to Weinstein and Yildiz (2007), is that each action is strictly dominant for some state of nature:

[8] As is standard, the Cartesian product of a collection of topological spaces $(V_\lambda)_{\lambda\in\Lambda}$ is denoted by V, with typical element v. Given $\lambda \in \Lambda$, we write $V_{-\lambda}$ for $\prod_{\ell\in\Lambda\setminus\{\lambda\}} V_\ell$, with typical element $v_{-\lambda}$. Likewise, given a family $g_\lambda : Y_\lambda \to Z_\lambda$ of functions, we write $g(y)$ and $g_{-\lambda}(y_{-\lambda})$ for $(g_\lambda(y_\lambda))_{\lambda\in\Lambda}$ and $(g_{\lambda'}(y_{\lambda'}))_{\lambda'\neq\lambda}$, respectively.

Assumption R-Dom (Richness-Dominance (Weinstein and Yildiz, 2007, Ass. 1)). *For each player $i = 1, 2$ and each action $a_i \in A_i$, there is a state $\theta^{a_i} \in \Theta$ of nature such that*

$$u_i(a_i, a_{-i}, \theta^{a_i}) > u_i(a_i', a_{-i}, \theta^{a_i})$$

for all $a_i' \neq a_i$ and $a_{-i} \in A_{-i}$.

An alternative richness condition is that beliefs about payoffs do not fully determine play. That is, for some beliefs about nature, a player can have multiple (strict) best responses, depending on his conjecture about the play of his opponent.

Assumption R-Mult(A') (Richness-Multiplicity). *Given a product set $A' \subset A$ with $|A_i'| > 1$ for all i, for each player $i = 1, 2$, there is a belief $\mu_i \in \Delta(\Theta)$ such that*

1. *for each $a_i \in A_i'$, there is a measurable function $\widetilde{s}_{-i}^{a_i} : \Theta \to \Delta(A_{-i}')$ such that*

$$\int_\Theta u_i(a_i, \widetilde{s}_{-i}^{a_i}(\theta), \theta) d\mu_i(\theta) > \int_\Theta u_i(a_i', \widetilde{s}_{-i}^{a_i}(\theta), \theta) d\mu_i(\theta) \qquad \text{for } a_i' \neq a_i;$$

2. *if $a_i' \notin A_i'$, then there is no measurable function $\widetilde{s}_{-i}^{a_i'} : \Theta \to \Delta(A_{-i})$ such that*

$$\int_\Theta u_i(a_i', \widetilde{s}_{-i}^{a_i'}(\theta), \theta) d\mu_i(\theta) \geq \int_\Theta u_i(a_i'', \widetilde{s}_{-i}^{a_i'}(\theta), \theta) d\mu_i(\theta) \qquad \text{for } a_i'' \neq a_i'.$$

The set of such beliefs μ is denoted by $\Delta_i^{A'}$.

In words, if Assumption R-Mult(A') is satisfied for a product set A', then each player i has a first-order belief μ_i about payoffs such that any action in A_i' is a strict best response against some conjecture that the opponent plays an action in A_{-i}' and these are the only best responses. While Assumption R-Mult(A') is sufficient for our results to hold, we conjecture it is not necessary, just like Assumption R-Dom is not necessary for Weinstein and Yildiz's (2007) results (Penta, 2013).

Both richness conditions are satisfied if the set of possible payoff functions is sufficiently rich. For example, if the set of possible payoff functions includes all functions (i.e., $\Theta := [0, 1]^A \times [0, 1]^A$, and $u_i(a, \theta) := \theta_i(a)$), then both conditions are satisfied. Global games (e.g., Examples 1 and Section 5) also satisfy both conditions (with $A' = A$).[9]

[9] The two richness conditions are independent: Assumption R-Mult(A') does not imply Assumption R-Dom or vice versa. For example, a complete-information game (i.e., $\Theta = \{\theta\}$) with multiple strict equilibria obviously satisfies R-Mult(A') but does not satisfy R-Dom. A simple example that satisfies R-Dom but not R-Mult(A') is one where there are two states, θ_1, θ_2, and two actions, a_i^1, a_i^2, for each player, where each player i receives 1 if he plays a_i^ℓ in θ_ℓ and 0 otherwise.

3.3 Beliefs

We are interested in how higher-order beliefs impact strategic behavior. Higher-order beliefs can be modeled using belief hierarchies, where a belief hierarchy for player i specifies his belief about the state of nature and the other players' signals (i.e., about $\Theta \times X_{-i}$), his beliefs about his opponent's beliefs, and so on, up to some order.[10] We allow for an arbitrary depth of reasoning: a belief hierarchy can have any finite or infinite depth. To consider all possible specifications of players' higher-order beliefs, we construct the space of all belief hierarchies. To that aim, we construct two sequences of spaces for each player i, H_i^m and $\widetilde{H}_i^m$, $m \geq 0$, with H_i^m the set of *mth-order belief hierarchies* that "stop" reasoning at order m, and $\widetilde{H}_i^m$ the set of belief hierarchies that "continue" to reason at that order. The belief hierarchies in H_i^m are precisely the belief hierarchies that have depth m, while the belief hierarchies in $\widetilde{H}_i^m$ are used to construct the belief hierarchies that have depth at least $m+1$ (possibly infinite).

It will be convenient to fix two (arbitrary) labels $h_i^{*,0}$ and $\tilde{\mu}_i^0$. The label $h_i^{*,0}$ is similar to the level-0 type in the level-k literature. The label $\tilde{\mu}_i^0$ is a notational placeholder that will be used to construct other types. Let $\widetilde{H}_i^0 := X_i \times \{\tilde{\mu}_i^0\}$ and $H_i^0 := X_i \times \{h_i^{*,0}\}$ be the set of zeroth-order belief hierarchies that continue and that stop at order 0, respectively.[11]

We next consider players' beliefs about the state of nature and about whether the other players have stopped reasoning at order 0. Let

$$\widetilde{\Omega}_i^0 := \Theta \times \left(\widetilde{H}_{-i}^0 \cup H_{-i}^0\right);$$
$$\Omega_i^0 := \Theta \times H_{-i}^0.$$

Define the set of first-order belief hierarchies that continue and stop at order 1 by,

$$\widetilde{H}_i^1 := \widetilde{H}_i^0 \times \Delta\left(\widetilde{\Omega}_i^0\right);$$
$$H_i^1 := \widetilde{H}_i^0 \times \Delta\left(\Omega_i^0\right);$$

respectively. These equations describe the first-order beliefs for belief hierarchies that reason beyond the first order and that stop reasoning at the first order, respectively (where a first-order belief describes a player's belief about the state of nature and other player's signal). The first-order belief hierarchies in $\widetilde{H}_i^1$ will be used to define types of depth greater than 1, while the first-order belief hierarchies in H_i^1 define the depth-1 types. Types in $\widetilde{H}_i^1$ thus think

[10] We thus distinguish between a player's private information (i.e., signal) and his belief hierarchy/type, as is common in the literature on the robustness of game-theoretic predictions (e.g., Bergemann and Morris, 2005).

[11] The types in $\widetilde{H}_i^0$ and H_i^0 are introduced merely for notational convenience. Alternatively, we could have started with two copies of $\Delta(\Theta)$: one to describe the first-order beliefs of depth-1 types, and one to describe the first-order beliefs of types that have depth greater than 1.

possible that the other player has not yet stopped reasoning; this will allow us to model that a type of depth $k > 1$ thinks that the opponent has depth m at least 1.

We now define inductively the sets of higher-order belief hierarchies. For $k = 1, 2, \ldots$, suppose that for each player j and all $\ell \leq k$, $\widetilde{H}_j^\ell$ and H_j^ℓ are the sets of belief hierarchies that continue to reason beyond order ℓ and that stop reasoning at that order, respectively. Define

$$\begin{aligned} \widetilde{H}_i^{\leq k} &:= \widetilde{H}_i^k \cup \textstyle\bigcup_{\ell=0}^k H_i^\ell, & \widetilde{\Omega}_i^k &:= \Theta \times \widetilde{H}_{-i}^{\leq k}, \\ H_i^{\leq k} &:= \textstyle\bigcup_{\ell=0}^k H_i^\ell, & \Omega_i^k &:= \Theta \times H_{-i}^{\leq k}, \end{aligned}$$

and let

$$\widetilde{H}_i^{k+1} := \left\{(x_i, \mu_i^0, \ldots, \mu_i^k, \mu_i^{k+1}) \in \widetilde{H}_i^k \times \Delta(\widetilde{\Omega}_i^k) : \mathrm{marg}_{\widetilde{\Omega}_i^{k-1}} \mu_i^{k+1} = \mu_i^k\right\}, \tag{3.1}$$

$$H_i^{k+1} := \left\{(x_i, \mu_i^0, \ldots, \mu_i^k, \mu_i^{k+1}) \in \widetilde{H}_i^k \times \Delta(\Omega_i^k) : \mathrm{marg}_{\widetilde{\Omega}_i^{k-1}} \mu_i^{k+1} = \mu_i^k\right\}. \tag{3.2}$$

Again, the interpretation is that $\widetilde{H}_i^{k+1}$ is the set of belief hierarchies that continue to reason at order $k+1$, while the set H_i^{k+1} contains the hierarchies that stop reasoning at $k+1$. As before, the former can conceive of the possibility that the other players have not stopped reasoning at order k, while the latter cannot. A belief hierarchy $h_i^k \in H_i^k$ that stops reasoning at order k is said to have *depth (of reasoning)* k. The condition on the marginal in (3.1) and (3.2) is a standard coherency condition: it ensures that the beliefs at different orders do not contradict each other (see, e.g., Dekel and Siniscalchi, 2014, for a discussion). Define

$$H_i^\infty := \left\{\left(x_i, \mu_i^0, \mu_i^1, \ldots\right) : \left(x_i, \mu_i^0, \ldots, \mu_i^k\right) \in \widetilde{H}_i^k \text{ for all } k \geq 0\right\}.$$

The belief hierarchies in H_i^∞ are those that "reason up to infinity." We therefore say that a belief hierarchy h_i^∞ in H_i^∞ has an *infinite depth (of reasoning)*. The set of all belief hierarchies is thus

$$H_i := H_i^\infty \cup \bigcup_{k=0}^\infty H_i^k.$$

With some abuse of notation, we sometimes write $(x_i, \mu_i^0, \ldots)$ for an element h_i of H_i, regardless of whether h_i has finite or infinite depth.

3.4 Universal type space

Following Mertens and Zamir (1985), we can use the sets H_i of all belief hierarchies to define the universal type space. A key observation is that every belief hierarchy h_i corresponds to a belief $\psi_i(h_i) \in \Delta(\Theta \times H_{-i})$ about nature and other players' hierarchies.

Proposition 3.1. *There is unique mapping $\psi_i : H_i \to \{h_i^{*,0}\} \cup \Delta(\Theta \times H_{-i})$ with the property that for each $k = 1, 2, \ldots$, for each $h_i = (x_i, \mu_i^1, \ldots) \in H_i$ of depth at least k, its kth-order belief μ_i^k is given by*

$$\mu_i^k = marg_{\widetilde{\Omega}_i^{k-1}} \psi_i(h_i),$$

and is such that

- *for $h_i^0 \in H_i^0$, $\psi_i(h_i^0) = h_i^{*,0}$;*
- *for $h_i^k \in H_i^k$, $k < \infty$, the support of $\psi_i(h_i^k)$ lies in $\Theta \times H_{-i}^{\leq k-1}$;*
- *for $h_i^\infty \in H_i^\infty$, the support of $\psi_i(h_i^\infty)$ lies in $\Theta \times H_{-i}$.*

Moreover, the function ψ_i is continuous.

The result follows from Proposition B.3 in Appendix B. There, we show that the tuple $(H_i, \psi_i)_{i=1,2}$ defines a type space, denoted $\mathcal{T}^*$, with H_i a set of *types* and for each type $h_i \in H_i$ a belief $\psi_i(h_i) \in \Delta(\Theta \times H_{-i})$ over the payoff parameters and the other player's types. As we show, $\mathcal{T}^*$ is universal in the sense that it generates all belief hierarchies with a finite or infinite depth. For simplicity, we sometimes write ψ_{h_i} for the belief $\psi(h_i)$ associated with the type h_i.

In some instances, a researcher may want to rule out certain beliefs. For example, in an auction setting, he may want to assume that a player's expected valuation increases in his signal. This can be captured using models:

Definition 1. A *model* is a pair $M = (\widetilde{\Theta}, T)$, where $\widetilde{\Theta} \subset \Theta$ and $T := (T_i)_{i=1,2}$ is a pair of subsets $T_i \subset H_i$ of types such that for each type $h_i \in T_i \setminus H_i^0$, ψ_{h_i} has support in $\widetilde{\Theta} \times T_{-i}$. A model is *finite* if T is finite.

Note that whether or not a model is finite is unrelated to the depth of reasoning of the types. For example, a finite model may include only infinite-depth types, and a model that is not finite may consist of types of any (finite or infinite) depth.

3.5 (Almost) common belief in an infinite depth

We are interested in relaxing the standard assumptions on beliefs about players' depth of reasoning. We first make these assumptions explicit in the context of the universal type space $\mathcal{T}^*$. Define

$$C_i^1 := \left\{ h_i \in H_i^\infty : \psi_{h_i}\left(\Theta \times H_{-i}^\infty\right) = 1 \right\}$$

to be the set of types that have an infinite depth of reasoning that believe that the opponent has an infinite depth. Since types from standard (Harsanyi) type spaces all generate belief

hierarchies with an infinite depth, standard types all satisfy this assumption. For $n > 1$, let

$$C_i^n := \Big\{ h_i \in C_i^{n-1} : \psi_{h_i}\Big(\Theta \times C_{-i}^{n-1}\Big) = 1 \Big\}$$

be the set of infinite-depth types that have nth-order mutual belief in an infinite depth. Again, all types from standard type spaces satisfy this condition. Let $C_i^\infty := \bigcap_n C_i^n$. Then, $C^\infty := (C_i^\infty)_{i=1,2}$ is the event in $\mathcal{T}^*$ that types have an infinite depth of reasoning and there is *common (correct) belief in the event that players have an infinite depth of reasoning.* It is easy to see that C^∞ is a model. The next result states that the types from standard type spaces are precisely the types in C^∞.

Proposition 3.2. [Common belief in infinite depth] *The universal type space $\mathcal{T}^{MZ}$ of Mertens and Zamir (1985) corresponds to the event in $\mathcal{T}^*$ that players have an infinite depth of reasoning and this is common belief: there is a belief-preserving homeomorphism from $\mathcal{T}^{MZ}$ to C^∞.*

See the online appendix for a formal statement and the proof. With this characterization in hand, we can weaken the strong assumption that there is common belief in an infinite depth by allowing for small deviations of this assumption. We thus consider the event that players have an infinite depth of reasoning and this almost-common belief. To construct this event, we define a sequence $B_i^1, B_i^2, \ldots$ of subsets of types. For each player i, let

$$B_i^0 := \Big\{ h_i \in H_i^\infty : \psi_{h_i}\Big(\bigcup_{\gamma<\infty} H_{-i}^\gamma\Big) = 1 \Big\}$$

be the set of types that have an infinite depth and that believe (with probability 1) that the other player has a finite depth. For $n = 1, 2, \ldots$, define

$$B_i^n := \Big\{ h_i \in H_i^\infty : \psi_{h_i}\Big(B_{-i}^{n-1}\Big) = 1 \Big\}.$$

Thus, if player i has a type in B_i^1, then he has an infinite depth of reasoning, and he believes that his opponent has an infinite depth but that she believes that he has a finite depth. Generally, the types in B_i^n have an infinite depth of reasoning and have nth-order mutual belief in the event that players have an infinite depth of reasoning, but not $(n+1)$th-order mutual belief in that event.

We will also consider the analogous conditions under the assumption that players have *level-k beliefs*. Level-k beliefs have been explored in the experimental literature. The assumption is that players with finite depth ℓ believe that their opponents have depth $\ell - 1$. For $\ell = 1, 2, \ldots$, let $G_i^\ell \subset H_i^\ell$ be the set of types for player i that believe that his opponent has depth $\ell - 1$,

believe that his opponent believes that he has depth $\ell - 1 - 1$, and so on (i.e., $\psi_{h_i}(G_{-i}^{\ell-1}) = 1$ for $h_i \in G_i^{\ell}$). Define

$$\widetilde{B}_i^0 := \Big\{ h_i \in H_i^\infty : \psi_{h_i}\Big(\bigcup_{\gamma<\infty} G_{-i}^{\gamma} \Big) = 1 \Big\},$$

and for $n = 1, 2, \ldots$, define $\widetilde{B}_i^n$ analogously to B_i^n. Clearly, $\widetilde{B}^n \subset B^n$. Like $B^0, B^1, \ldots$, the sequence $\widetilde{B}^0, \widetilde{B}^1, \ldots$ captures that players have an infinite depth of reasoning and have almost common belief in the event that players have an infinite depth of reasoning; in addition, the only finite-depth types that types in $\widetilde{B}^n$ think possible have level-k beliefs.

3.6 Rationalizability

We next extend the standard definition of rationalizability to our environment. As we show in Appendix A, it suffices to define rationalizability for the universal type space $\mathcal{T}^*$ since the set of rationalizable actions of a type depends only on its induced belief hierarchy, as in the standard case (Dekel, Fudenberg, and Morris, 2007). For each player $i = 1, 2$ and $h_i \in H_i \setminus H_i^0$, and define

$$R_i^0(h_i) := A_i.$$

For $m > 1$, define inductively[12]

$$R_i^m(h_i) := \left\{ a_i \in A_i : \begin{array}{l} \text{there is a measurable } s_{-i} : \Theta \times H_{-i} \to \Delta(A_{-i}) \text{ s.t.} \\ \text{supp } s_{-i}(\theta, h_{-i}) \subseteq R_{-i}^{m-1}(h_{-i}) \text{ for all } h_{-i} \in H_{-i}, \theta \in \Theta; \text{ and} \\ a_i \in \arg\max_{a_i' \in A_i} \int_{\Theta \times H_{-i} \times A_{-i}} u_i(a_i', a_{-i}, \theta) s_{-i}(\theta, h_{-i})(a_{-i}) \mathrm{d}\psi_{h_i} \end{array} \right\}.$$

where supp μ is the support of a probability measure μ. Then, $R_i^m(h_i)$ is the set of *m-rationalizable actions* for h_i. The m-rationalizable actions for a type are the actions that survive m rounds of iterated deletion of dominated actions: for each action $a_i \in R_i^m(h_i)$, there is a conjecture s_{-i} that rationalizes it, in the sense that the conjecture has support in the actions of the opponents that have survived $m-1$ rounds of deletion, and a_i is a best response to this conjecture (given the type's belief ψ_{h_i}). The *(interim correlated) rationalizable actions* of type h_i are

$$R_i^\infty(h_i) := \bigcap_{m=0}^{\infty} R_i^m(h_i).$$

If $h_i \in H_i^0$, then we set $R_i^\infty(h_i) := A_i$.

For types that have an infinite depth of reasoning and common belief in this event, interim correlated rationalizability captures the behavioral implications of rationality and common

[12]For notational convenience, the conjecture s_{-i} in the definition of $R_i^m(h_i)$ is defined also for types h_{-i} outside the support of ψ_{h_i}. This does not affect our results.

belief in rationality (Dekel, Fudenberg, and Morris, 2007). If a types has finite depth k, then its set of rationalizable actions is completely determined by the actions that survive k rounds of elimination:

Lemma 3.3. *Let h_i be a type with finite depth k. Then, an action a_i is k-rationalizable for h_i if and only if it is rationalizable for the type.*

In Appendix C, we show that rationalizability satisfies the standard best-reply property: any rationalizable action for a type is a best response to the belief that the opponent chooses a rationalizable strategy (cf. Dekel, Fudenberg, and Morris, 2007, Prop. 4).

The rationalizability correspondence satisfies a standard upper hemicontinuity property (cf. Dekel, Fudenberg, and Morris, 2007, Lemma 1):

Lemma 3.4. [Upper hemicontinuity] *The rationalizability correspondence is nonempty and upper hemicontinuous: every type $h_i \in H_i$ has a neighborhood $O(h_i)$ such that $R_i^\infty(h_i) \supset R_i^\infty(h_i') \neq \emptyset$ for $h_i' \in O_i(h_i)$. Likewise, the m-rationalizability correspondence is nonempty and upper hemicontinuous.*

The result says that if a player's actual type is h_i but a researcher has only an imperfect observation of the player's type (i.e., he observes that the type is in $O_i(h_i)$), then any action that is rationalizable for a type that is consistent with his observation (i.e., $a_i \in R_i^\infty(h_i')$ for $h_i' \in O_i(h_i)$) is rationalizable for the player's actual type (i.e., $a_i \in R_i^\infty(h_i)$), and likewise for finite-order rationalizability.

A researcher who cannot perfectly observe the players' types may want his predictions to be robust to small perturbations in beliefs. The concept of robust rationalizability captures such a robustness requirement:

Definition 2. A subset $A_i' \subset A_i$ of actions is *robustly rationalizable* for a type h_i if the rationalizable actions for h_i are precisely the actions in A_i' and R_i^∞ is locally constant at h_i: h_i has a neighborhood $O_i(h_i)$ such that $R_i^\infty(h_i') = A_i'$ for every $h_i' \in O_i(h_i)$. Likewise, A_i' is *robustly m-rationalizable* for h_i if the m-rationalizable actions for h_i are precisely the actions in A_i' and R_i^m is locally constant at h_i.

A direct corollary of Lemma 3.4 is that uniqueness is robust:

Corollary 3.5. [Uniqueness robust] *If action a_i is the unique rationalizable action for a type h_i, then $\{a_i\}$ is robustly rationalizable for h_i.*

Weinstein and Yildiz (2007) show the surprising result that if players have an infinite depth and this is common belief, then *only* uniqueness is robust:

Proposition 3.6. [No robust multiplicity with common belief in an infinite depth; Weinstein and Yildiz, 2007, Prop. 2] *Under Assumption R-Dom, A'_i is robustly rationalizable for a type h_i with common belief in an infinite depth (i.e., $h_i \in C_i^\infty$) only if $A'_i = \{a_i\}$.*

Thus, if a researcher thinks that players have an infinite depth of reasoning and this is common belief, then the only robust predictions he can make is that players have a unique rationalizable action. In the next section, we show that this extreme conclusion is not robust to relaxing the assumption that there is common belief in an infinite depth of reasoning.

4 Robust multiplicity

In this section we show that the rationalizable actions of a type h_i with multiple rationalizable actions can be robust to perturbations of beliefs at h_i when there is uncertainty about players' depth of reasoning. Section 4.1 considers the general case to derive basic insights on the rationalizability correspondence. Section 4.2 focuses on the case of complete-information types to show that, unlike in the standard case with common belief in an infinite depth, the rationalizability correspondence is continuous around complete-information types when we relax standard assumptions. Section 5 builds on the insights developed in this section to develop robust refinements.

4.1 General

We first consider general types. The main result of this section shows that if we relax the assumption that it is common belief that players have an infinite depth of reasoning, then multiplicity can be robust.

Theorem 4.1. [Robust multiplicity with almost-common belief in infinite depth] *Under Assumption R-Mult(A'), for every $m = 1, 2, \ldots$, there is an infinite-depth type h_i^m with mth-order mutual belief in an infinite depth for whom A'_i is robustly rationalizable.*

Proof. For simplicity, we say that a type h_i has a belief in the multiplicity set if $\text{marg}_\Theta \psi_{h_i} \in \Delta_i^{A'}$, and we write $h_i \in \Delta_i^{A'}$. The proof has two key ingredients. The first provides a "grain" of robust multiplicity:

Lemma 4.2. *Under Assumption R-Mult(A'), the set $\{h_i \in H_i : h_i \in \Delta_i^{A'}\}$ is nonempty and open.*

The set in Lemma 4.2 is a "grain" of robust multiplicity in the sense that it almost immediately defines a set of depth-1 types with robust multiplicity, as we show below. The second key

ingredient provides us with a contagion-type argument: given a grain V of robust multiplicity, we can identify other types with robust multiplicity.

Lemma 4.3. *Under Assumption R-Mult(A'), for $m = 0, 1, 2, \ldots, \infty$, if $V \subset H_{-i}$ is an open subset of types such that $R^m_{-i}(h_{-i}) = A'_{-i}$ for $h_{-i} \in V$, then every type h_i with a belief in the multiplicity set that assigns probability 1 to V has a neighborhood $O(h_i)$ such that $R^{m+1}_i(h'_i) = A'_i$ for all $h'_i \in O(h_i)$ (where $\infty + 1 = \infty$).*

Lemma 4.3 says that if there is an open set V of types for whom the the set of rationalizable actions is A', then A' is robustly rationalizable for any type that has a belief in the multiplicity set and that believes that the opponent has a type in V, and similarly for finite-order rationalizability.

With these tools in hand, we can prove Theorem 4.1. The first step is to show that multiplicity is robustly m-rationalizable for any finite m:

Lemma 4.4. *Under Assumption R-Mult(A'), for every $m = 1, 2, \ldots$, there are types for whom the actions in A'_i are robustly m-rationalizable.*

The proof of Lemma 4.4 can be found in the appendix. The proof uses Lemma 4.2 to show that the set of types for whom A' is 1-rationalizable is open. This gives a set of types for whom A' is robustly 1-rationalizable. The proof then applies Lemma 4.3 repeatedly to identify sets of types for whom A' is robustly m-rationalizable.

Together with Lemma 3.3, Lemma 4.4 immediately implies that multiplicity is robust for finite-depth types:

Corollary 4.5. *Under Assumption R-Mult(A'), for every $m = 1, 2, \ldots$, there are depth-m types for whom A'_i is robustly rationalizable. In fact, for every $m = 1, 2, \ldots$, there is an open set $V^m_i \subset H^m_i$ of depth-m types such that $R^\infty_i(h_i) = A'_i$ for $h_i \in V^m_i$.*

The second claim in Corollary 4.5 follows immediately from the first: by the first claim, there is an open set O^m_i of types for whom A'_i is robustly rationalizable that satisfies $O^m_i \cap H^m_i \neq \emptyset$. The second claim then follows by taking $V^m_i := O^m_i \cap H^m_i$. It is straightforward to use the contagion argument in Lemma 4.3 to show that multiplicity can be robust for types with an infinite depth that have almost-common belief in an infinite depth; see the appendix. □

The contrast between the negative result for the case with common belief in an infinite depth (Proposition 3.6) and the positive result for the case with almost-common belief in an infinite depth (Theorem 4.1) is stark: if the set of possible payoff functions is sufficiently rich (i.e., satisfies Assumptions R-Dom and R-Mult(A')), then multiplicity is not robust when it is

common belief that players have an infinite depth of reasoning, yet it can be robust if there is only almost-common belief in an infinite depth of reasoning.

Theorem 4.1 implies that the structure of the correspondence that maps beliefs into sets of rationalizable actions changes significantly if we relax the assumption that it is common belief that players have an infinite depth. If there is common belief in an infinite depth, then types generically have a unique rationalizable action, that is, the set of types with a unique rationalizable action is open and dense in the set of types with common belief in an infinite depth (Weinstein and Yildiz, 2007). Theorem 4.1 implies that if we weaken the assumption that players have common belief in an infinite depth, then uniqueness is no longer generic.[13]

The deviation from the standard assumption that there is common belief in an infinite depth is arguably minimal: the types in Theorem 4.1 are arbitrarily close to standard types. We could even make more stringent assumptions on players' beliefs about reasoning ability. For example, the result holds also if we require that players have level-k beliefs (i.e., if we replace B^n by $\widetilde{B}^n$ throughout) or require that infinite-depth types assign high probability only to types with a high depth of reasoning. In fact, the same result obtains in a universal type space in which every type has depth at least k for arbitrary finite k.

It is worth asking where the proof of Theorem 4.1 breaks down if there is common belief that players have an infinite depth. The central contagion argument, Lemma 4.3, goes through also if there is common belief in an infinite depth. So, the analogue of Lemma 4.4 for $\mathcal{T}^{MZ}$ also holds: under Assumption R-Mult(A'), for every $m = 1, 2, \ldots$, there are types in the universal space $\mathcal{T}^{MZ}$ for standard types for whom the actions in A' are robustly m-rationalizable. However, by the results of Weinstein and Yildiz (2007), there is no type in $\mathcal{T}^{MZ}$ for whom the actions in A' are robustly rationalizable.[14] This reflects the well-known extreme sensitivity of the rationalizability correspondence to tail events in the standard case with common belief in an infinite depth (e.g., Rubinstein, 1989; Carlsson and van Damme, 1993; Weinstein and Yildiz, 2007, amongst many others).

The behavior of the rationalizability correspondence is thus very different depending on whether there is common belief in an infinite depth or almost-common belief in an infinite

[13]Formally, Theorem 4.1 shows that the set of types with multiple rationalizable actions has a nonempty interior in the universal type space $\mathcal{T}^*$. This implies that the set of types with a unique rationalizable actions is not generic (i.e., is not open and dense). However, Theorem 4.1 is strictly stronger. For example, the set $\mathbb{Q}$ of rationals is not open and dense in the set $\mathbb{R}$ of real numbers, yet the interior of its complement $\mathbb{R} \setminus \mathbb{Q}$ is empty.

[14]Formally, if we define S^m and $S^{MZ,m}$ to be the set of types in $\mathcal{T}^*$ and $\mathcal{T}^{MZ}$, respectively, for whom the actions in A' are m-rationalizable, then $\bigcap_m S^{MZ,m}$ has empty interior (even though the intersection itself is nonempty) while $\bigcap_m S^m$ has a nonempty interior. One might think that the interior will be empty if we perform enough rounds of elimination for the types in $\mathcal{T}^*$. This is not the case; see Appendix C.

depth. This can be seen most clearly by comparing the behavior of the rationalizability correspondence around a standard type (with common belief in an infinite depth) with its behavior around a type with almost-common belief in an infinite-depth that has the same rationalizable actions. Consider a type $h_i^{MZ} \in C_i^\infty$ with common belief in an infinite depth that has multiple rationalizable actions. Since action sets are finite, there is some finite k such that $R_i^\infty(h_i^{MZ}) = R_i^k(h_i^{MZ})$. Then, there is a type $h_i \notin C_i^\infty$ that has the same ℓth-order belief hierarchy as h_i for some $\ell \geq k$ and thus $R_i^\infty(h_i) = R_i^\infty(h_i^{MZ})$ (Lemma A.1). In fact, we can take h_i to be consistent with mth-order mutual belief in an infinite depth for some m (i.e., $h_i \in B_i^m$). While multiplicity is not robust for h_i^{MZ} (by Proposition 3.6), it *is* robust for h_i (by Theorem 4.1).[15] That multiplicity is not robust for h_i^{MZ} just reflects the sensitivity of the rationalizability correspondence to tail events when there is common belief in an infinite depth: for every neighborhood $O(h_i^{MZ})$ of h_i^{MZ} and every ℓ, there is a type in $O(h_i^{MZ})$ for whom the set of rationalizable actions is sensitive to perturbations of the beliefs at order ℓ. By contrast, the rationalizability correspondence around h_i is *finitely determined* in the following sense: there is a neighborhood $O(h_i)$ of h_i and a finite order n such that for all types in $O(h_i)$, the set of rationalizable actions is not sensitive to changes in beliefs at orders greater than n.[16] To see the intuition, consider an infinite-depth type $h_i' \in B_i^0$ that assigns positive probability to types of depth m for every $m < \infty$. Then, for increasing k, its belief $\psi_{h_i'}$ puts vanishingly small weight on types with depth greater than k. So, while we can perturb its beliefs at arbitrarily high orders, the effect of perturbations of high-order beliefs on behavior will be small. Likewise, for any type $h_{-i}'' \in B_{-i}^1$ that puts probability 1 on h_i', the effect of belief perturbations at high order is limited. And so on. This reveals a subtle way in which a small "grain of naiveté" can give rise to robust predictions: even if the set of rationalizable actions of a type itself is unaffected, it ensures that the type has a neighborhood on which the rationalizability correspondence is not sensitive to perturbations of high-order beliefs.

This example also illustrates that the robustness of multiplicity in our setting cannot be traced back directly to the robustness of multiplicity under finite-order rationalizability. While for every type $h_i' \in O(h_i)$, there is some finite $n_{h_i'}$ such that $R_i^\infty(h_i') = R_i^{n_{h_i}}(h_i')$ (since action sets are finite), the neighborhood $O(h_i)$ contains uncountably many types. So, it is not immediate from the relation between rationalizability and finite-order rationalizability that the rationalizability correspondence around h_i is finitely determined (i.e., that there is a finite n

[15] To be precise, the latter statement presumes that the beliefs of h_i^{MZ} and h_i are as in the proof of Theorem 4.1. We can always find types for which this is the case.

[16] This is true even if h_i is arbitrarily far from the finite-depth types (i.e., m large). This result also does not hinge on types with a limited reasoning ability being close to h: we can choose the neighborhood in such a way that it includes only types with an infinite depth that assign high probability to types with a finite depth, and so on.

such that $n_{h_i'} \leq n$ for all $h_i' \in O(h_i)$).

In Section 5, we build on these insights to show that robust refinements of rationalizability exists. Before exploring this, we apply the insights developed in this section to show that there need not be a discontinuity of behavior around complete-information types when we relax the assumption that there is common belief in an infinite depth.

4.2 Almost-complete information

The same basic tools can be used to study robustness in specific applications. We focus on the case where payoffs are observed with some small noise. Again, consider the well-known example from Carlsson and van Damme (1993):

	I	NI
I	θ, θ	$\theta - 1, 0$
NI	$0, \theta - 1$	$0, 0$

where $\theta \in \Theta := [-1, 2]$. Both Assumptions R-Dom and R-Mult(A') (with $A_i' = \{I, NI\}$) are satisfied: If $\theta > 1$, then investing (I) is strictly dominant for each player; if $\theta < 0$, then not investing (NI) is strictly dominant. For $\theta \in (0, 1)$, either action is a strict best response for a player depending on his conjecture about the opponent's play.

The information structure is as described in Example 1. So, each player i observes signal x_i of the state θ that is potentially noisy. The noise is measured by a parameter $\varepsilon \in [0, \frac{1}{2})$. Players' observations of θ become increasingly precise as the noise parameter ε goes to 0. It will be convenient to take the set of signals to be $X_i := [-2, 3]$. When the noise parameter is ε, the set of signals that types receive with positive probability is thus $X_i^\varepsilon := [-1 - \varepsilon, 2 + \varepsilon]$.

We consider a simple model with level-k beliefs that is consistent with this information structure. For each signal $x_i \in X_i^\varepsilon$, there is a unique type h_i^{1,x_i} in H_i^1 with signal x_i whose beliefs are consistent with the information structure. Let $T_i^{\varepsilon,1} := \{h_i^{0,x_i} : x_i \in X_i^\varepsilon\}$ be the set of such types. For $m > 1$, suppose $T_i^{\varepsilon,m-1}$ is a set of depth-$(m-1)$ types with beliefs consistent with the information structure such that for every $x_i \in X_i^\varepsilon$, there is a type in $T_i^{\varepsilon,m-1}$ with signal x_i. Then, for each signal $x_i \in X_i^\varepsilon$, there is a unique depth-m type h_i^{m,x_i} with signal x_i whose beliefs are consistent with the information structure and that assigns probability 1 to $T_{-i}^{\varepsilon,m-1}$. Let $T_i^{\varepsilon,m} := \{h_i^{m,x_i} : x_i \in X_i^\varepsilon\}$ be the set of such types. We next define the types with an infinite depth of reasoning. Fix some vector $(p_1, p_2, \ldots)$ of probabilities (i.e., $p_n \geq 0$ for all n, and $\sum_n p_n = 1$). Then, for each $x_i \in X_i^\varepsilon$, there is a unique infinite-depth type h_i^{∞,x_i} in $\widetilde{B}_i^0$ with signal x_i whose beliefs consistent with the information structure that assigns probability p_n to $T_{-i}^{\varepsilon,n}$, $n = 1, 2, \ldots$. Let $T_i^{\varepsilon,\infty} := \{h_i^{\infty,x_i} : x_i \in X_i^\varepsilon\}$. For $n > 0$, given a set $T_i^{\varepsilon,\infty+n-1} \subset \widetilde{B}_i^{n-1}$

of types with beliefs consistent with the information structure, we can likewise select, for each $x_i \in X_i^\varepsilon$, the unique type $h_i^{\infty+n,x_i}$ in $\widetilde{B}_i^n$ with signal x_i whose beliefs are consistent with the information structure; this defines a set $T_i^{\varepsilon,\infty+n}$ of types with an infinite depth and nth-order mutual belief in an infinite depth whose beliefs are consistent with the information structure. Let T_i^ε be the union of the sets $T_i^{\varepsilon,0}, T_i^{\varepsilon,1}, \ldots, T_i^{\varepsilon,\infty}, T_i^{\varepsilon,\infty+1}, \ldots$ (where $T_i^{\varepsilon,0} := X_i^\varepsilon \times \{h_i^{*,0}\}$). Then, $M^\varepsilon = (\Theta, T^\varepsilon)$ is a model with level-k beliefs that is consistent with the information structure. Moreover, it is consistent with almost-common belief in an infinite depth (i.e., for all n, $\widetilde{B}^n \cap T^\varepsilon \neq \emptyset$).

A first observation is that for any signal $x_i \in (0,1)$, both actions are rationalizable for finite-depth types in M^ε with signal x_i, provided that the noise is sufficiently small:

Lemma 4.6. *For every k, there exist $\underline{x}_k^\varepsilon \in (0, \frac{1}{2})$, and $\bar{x}_k^\varepsilon \in (\frac{1}{2}, 1)$ such that for every depth-k type h_i in T^ε, both actions are robustly rationalizable whenever its signal x_i is in $(\underline{x}_k^\varepsilon, \bar{x}_k^\varepsilon)$. Moreover, for every k, $\underline{x}_k^\varepsilon \to 0$ and $\bar{x}_k^\varepsilon \to 1$ as $\varepsilon \to 0$.*

The proof gives an explicit expression for the interval bounds $\underline{x}_k^\varepsilon, \bar{x}_k^\varepsilon$. The next result extends this to the case where if types have an infinite depth and have almost-common belief in infinite depth.

Theorem 4.7. [Robust multiplicity around complete-information types] *For every m, there exist $\varepsilon_m > 0$, $\underline{x}_m^\varepsilon \in (0, \frac{1}{2})$, and $\bar{x}_m^\varepsilon \in (\frac{1}{2}, 1)$ such that both actions are robustly rationalizable for any infinite-depth type $h_i \in T^\varepsilon$ with signal $x_i \in (\underline{x}_m^\varepsilon, \bar{x}_m^\varepsilon)$ and mth-order belief in an infinite depth whenever $\varepsilon < \varepsilon_m$. Moreover, we can choose the bounds $\underline{x}_m^\varepsilon$ and $\bar{x}^\varepsilon$ such that $\underline{x}_m^\varepsilon \to 0$ and $\bar{x}_m^\varepsilon \to 1$ as $\varepsilon \to 0$.*

Theorem 4.7 implies that the strategic discontinuity around complete-information games vanishes if we relax the assumption that there is common belief in an infinite depth of reasoning and instead require only that players have an infinite depth and this is almost-common belief. As the noise ε goes to 0, players can rationally choose both actions when neither action is dominant, just like in the limit case with complete information (i.e., $\varepsilon = 0$). So, there is no discontinuity when ε goes to 0. In particular, the risk-dominant strategy is not uniquely selected when ε is small but positive, unlike in the standard case (Carlsson and van Damme, 1993).

Importantly, the conclusions in Theorem 4.7 are robust to further belief perturbations. Since multiplicity is robust, the predictions remain valid even if the researcher has misspecified the model in the sense that the model T^ε does not accurately describe players' beliefs, as long as the assumptions embodied in the model are satisfied approximately.

The proof of Theorem 4.7 is similar in nature to that of Theorem 4.1: the key ingredients are a "grain" of robust multiplicity, provided by the finite-depth types, and a contagion argument

to establish robust multiplicity for types consistent with almost-common belief in infinite depth. The proofs differ slightly in that the contagion argument for Theorem 4.7 has to ensure that beliefs are consistent with the information structure. The proof of Theorem 4.7 thus illustrates how the basic contagion argument can be adapted to prove results for specific applications.

Again, Theorem 4.7 does not rely on strong assumptions on players' beliefs. First, it does not require that players believe that other players have a shallow depth of reasoning, or that others believe that, believe that others believe that, and so on. Indeed, the same result obtains if we assume that all players have depth at least k, for arbitrary finite k. Second, while T^ε is a minimal model consistent with the information structure and almost-common belief in an infinite depth (assuming level-k beliefs), the main insight extends to a much broader range of situations. In the online appendix, we show that multiplicity is robust for all types whose beliefs are consistent with the information structure and almost-common belief in an infinite depth, though the bound on the noise level may depend on the detailed features of a type in that case.

5 Robust refinements

In this section, we study the implications of robust multiplicity for the scope for robust refinements. We show that, unlike in the standard case with common belief in an infinite depth, robust refinements do exist when there is almost-common belief in an infinite depth.

5.1 Definitions

We start with defining some basic concepts. A strategy for a model $M = (\widetilde{\Theta}, T)$ is a measurable function σ_i that maps each type $h_i \in T_i$ into a mixed action $\sigma_i(h_i) \in \Delta(A_i)$. We write $\sigma_i(a_i \mid h_i)$ for the probability that i plays a_i when his type is h_i. A strategy profile $\sigma = (\sigma_i)_{i=1,2}$, with σ_i a strategy for M, is a *(Bayesian-Nash) equilibrium* for M if for each player $i = 1, 2$, $h_i \in T_i \setminus H_i^0$, $a_i \in A_i$ such that $\sigma_i(a_i \mid h_i) > 0$,[17]

$$\int u_i(a_i, \sigma_{-i}(h_{-i}), \theta) \mathrm{d}\psi_{h_i} \geq \int u_i(a_i', \sigma_{-i}(h_{-i}), \theta) \mathrm{d}\psi_{h_i}$$

for all $a_i' \in A_i$. As before, a nonstrategic type $h_i \in H_i^0$ can play any action. An equilibrium is *strict* if the above inequality is strict whenever $a_i' \neq a_i$. Bayesian-Nash equilibrium refines rationalizability: if σ is a Bayesian-Nash equilibrium for M, then every action a_i that is played

[17]This definition is consistent with the definition of Nash equilibrium in Strzalecki (2014) for complete-information games.

with positive probability by a type h_i in M under σ (i.e., $\sigma_i(a_i \mid h_i) > 0$) is rationalizable for the type (i.e., $a_i \in R_i^\infty(h_i)$).

To assess the robustness of equilibrium predictions, we again take the perspective of a researcher who can observe finitely many orders of beliefs with some noise. So, there is some finite order κ and some $\eta > 0$ such that for each $\ell < \kappa$, the researcher cannot rule out ℓth-order beliefs that are within η of the observed ℓth-order belief (in the usual weak topology).[18] In particular, if the researcher observes that a player has an mth-order belief, then he rules out that the player has a depth of reasoning strictly less than m.[19] We write $O_{\eta,\kappa}(t_i)$ for the set of types that the researcher finds possible if he observes the type t_i. We focus on the case where the noise is small, that is, $\eta \to 0$ and $\kappa \to \infty$.

Definition 3. A pair (M', τ) is an (η, κ)-*perturbation* of a model $M = (\widetilde{\Theta}, T)$ if $M' = (\widetilde{\Theta}', T')$ is a finite model and $\tau : T \to T'$ is such that $\tau_i(t_i) \in O_{\eta,\kappa}(t_i)$ for every $t_i \in T_i$.

This naturally leads to the following robustness requirement: since a researcher cannot rule out any (η, κ)-perturbation of a model, a prediction for a model is robust if it is valid for any (η, κ)-perturbation of the model. In the context of Bayesian-Nash equilibrium, this gives the following condition:[20]

Definition 4. A Bayesian-Nash equilibrium σ for $M = (\widetilde{\Theta}, T)$ is (η, κ)-*robust* if for every (η, κ)-perturbation (M', τ) of M, there is a Bayesian-Nash equilibrium σ' for M' that coincides with σ on $\tau(T)$ (i.e., for each player $i = 1, 2$, type $h_i \in T_i$, $\sigma'_i(\tau_i(h_i)) = \sigma_i(h_i)$). A Bayesian-Nash equilibrium is *robust* if it is (η, κ)-robust for some $\eta > 0$ and $\kappa < \infty$.

Equilibrium may provide stronger predictions than rationalizability in models with multiplicity, that is, models that have a type with multiple rationalizable actions. However, there may be a tension between the desire to get sharp prediction and the requirement that predictions be robust. The next result shows that if there is common belief in an infinite depth, then any equilibrium that makes stronger predictions than rationalizability is not robust:

[18] In addition, a researcher may want to measure a type's signal x_i. We abstract away from this for the purposes of this section; so, we will take $X_i = \{x_i\}$ in this section.

[19] More precisely, the topology on the set of mth-order beliefs is the usual weak topology. This topology can be metrized with the Prokhorov metric d_P^m; and we say that a type with mth-order belief μ_i^m is within η of a type with mth-order belief ν_i^m if $d_P^m(\mu_i^m, \nu_i^m) < \eta$. If η is not too large (e.g., $\eta < 1$), then the researcher can distinguish between types with different depths of reasoning $k, k' \leq \kappa$, as well as between a type that has depth $k \leq \kappa$ and a type with depth $k' > \kappa$.

[20] The robustness requirement in Definition 4 is stronger than the robustness requirement in Weinstein and Yildiz (2007). First, they do not require that the strategy profile for the perturbed model coincides exactly with the strategy profile on the original model, only that it produces the same behavioral patterns. Also, their definition of an (η, κ)-perturbation (M', τ) requires M' to admit a full support common prior. Working with a stronger robustness requirement strengthens our positive result.

Proposition 5.1. [No robust refinement under common belief in infinite depth; Weinstein and Yildiz, 2007] *Under Assumption R-Dom, if $M = (\widetilde{\Theta}, T)$ is a finite model with multiplicity that is consistent with common belief in an infinite depth (i.e., $T \subset C^{\infty}$) and $\widetilde{\Theta}$ is finite, then M does not have a robust equilibrium.*

So, under the assumption that there is common belief in an infinite depth, a researcher who is concerned with the robustness of his predictions cannot make sharper predictions by using a refinement of rationalizability. In the next section, we show that the situation is strikingly different when there is almost-common belief in an infinite depth.

5.2 Results

If we relax the assumption that there is common belief in an infinite depth, then robust refinements exist. To state the result, say that a *model has multiplicity* if it has a type with multiple rationalizable actions.

Theorem 5.2. [Robustness of strict equilibrium under almost-common belief in infinite depth] *Under R-Mult(A′), for every $m = 1, 2, \ldots$, there is a model M with multiplicity and mth-order belief in an infinite depth such that every strict Bayesian-Nash equilibrium for M is robust.*

Since equilibrium is a refinement of rationalizability and there are types with multiple rationalizable actions, Theorem 5.2 immediately implies that robust refinements of rationalizability exist.

Again, the contrast between the case with common belief in an infinite depth and almost-common belief in an infinite depth is stark. If there is common belief in an infinite depth, then the requirement that predictions be robust eliminates *all* equilibria if there is a type with multiple rationalizable actions (Proposition 5.1). In contrast, if we relax the assumption that there is common belief in an infinite depth, *every* strict equilibrium can be robust (Theorem 5.2). In that case, robustness does not impose any additional restrictions on equilibria beyond strictness.

Theorem 5.2 is not a simple corollary of the results for rationalizability. In the case of rationalizability, the key is to ensure that players can hold multiple conjectures about the opponent's behavior. In the case of equilibrium, players have a single conjecture that coincides with the opponent's actual strategy. The role of finite-depth types differs correspondingly. In the case of equilibrium, finite-depth types do not act as a grain of multiplicity, as in the case of rationalizability. Rather, their role is to "anchor" the behavior of high-depth types. Anchoring the behavior of types is critical: in environments with common belief in an infinite depth of

reasoning, the high-order beliefs of types can vary arbitrarily. But, if there is only almost-common belief in an infinite depth, then behavior can be determined at a finite order. To see this, consider an infinite-depth type that assigns high probability (say, $1-\eta$ for $\eta > 0$ small) to types of depth at most $k < \infty$. Then, since most of the probability mass is concentrated on types whose higher-order beliefs cannot be varied arbitrarily, perturbations of high-order beliefs have only a limited impact: the robustness of equilibrium behavior for the finite-depth types carries over to the infinite-depth types.

The proof of Theorem 5.2 illustrates the difference between the arguments for the case where types have a strictly dominant action and where they have multiple rationalizable actions. In the former case, strategic beliefs (i.e., beliefs about others' actions, beliefs about others' beliefs about actions, and so on) are irrelevant, and it follows directly that Bayesian-Nash equilibrium is robust to small perturbation of beliefs, even if there is common belief in an infinite depth. By contrast, if types have multiple best responses, a player's best response depends on his conjecture about the opponent's action. In this case, strategic beliefs matter and establishing robustness requires considerable care. In particular, it requires relaxing the assumption that there is common belief in an infinite depth so that finite-depth types, whose equilibrium actions do not depend on their high-order beliefs about strategies, can be used to "anchor" the behavior of types with a higher depth of reasoning.

5.3 Example: Global games

As noted above, robustness does not put much restrictions on equilibrium behavior. The global games introduced by Carlsson and van Damme (1993) provide a particularly clear illustration of this point. In global games, players play a supermodular game at each state θ. Each player i has two actions, labeled $\tilde{a}, \tilde{b}$. Each action $a_i = \tilde{a}, \tilde{b}$ is strictly dominant for player i at some state $\theta^{\tilde{a}_i}$; and for some θ^*, the complete-information game (with payoff functions $u_i(\cdot, \cdot, \theta^*) : A \to \mathbb{R}$) has two strict Nash equilibria, $(\tilde{a}, \tilde{a})$ and $(\tilde{b}, \tilde{b})$ (say). Clearly, every global game satisfies Assumptions R-Dom and R-Mult(A') (with $A'_i = \{\tilde{a}, \tilde{b}\}$). As in much of the applied literature, we assume that payoff functions are symmetric. Example 1 is an example of a symmetric global game.

The literature focuses on the question which equilibria of the complete-information game are robust to the introduction of a small amount of incomplete information about payoffs. The next result shows that if there is almost-common belief in an infinite depth, the requirement that predictions be robust does not rule out *any* of the equilibria of the complete-information game.

Proposition 5.3. [Robust equilibria in global games under almost-common belief

in an infinite depth] *Fix a symmetric global game. For every $m = 1, 2, \ldots$, there is a finite complete-information model M with mth-order belief in an infinite depth and robust equilibria $\sigma^{\tilde{a}}, \sigma^{\tilde{b}}$ for M such that if the complete-information game has multiple Nash equilibria (i.e., $\theta = \theta^*$), then players coordinate on $(\tilde{a}, \tilde{a})$ under $\sigma^{\tilde{a}}$, while they coordinate on $(\tilde{b}, \tilde{b})$ under $\sigma^{\tilde{b}}$.*

Proposition 5.3 shows that if we relax the strong assumption that there is common belief in an infinite depth, then every Nash equilibrium of the complete-information game is robust to the introduction of uncertainty about payoffs. This contrasts with the limit case with common belief in an infinite depth. In that case, if the complete-information game has multiple rationalizable actions, then the researcher is unable to make any robust predictions (Propositions 3.6 and 5.1).

6 Related literature

Almost-common belief in rationality. In the context of finitely repeated games, Kreps, Milgrom, Roberts, and Wilson (1982) and Milgrom and Roberts (1982) show that a grain of doubt about players' rationality can have a large impact on behavior. We likewise introduce a grain of doubt about the opponent's reasoning ability. However, there is a fundamental difference between their "irrational" types and our "naive," nonstrategic, types. While irrational types are (commonly known to be) committed to playing a certain action, we follow the level-k literature in assuming that nonstrategic types can play any action. Thus, in a sense, our approach can be viewed as relaxing common-knowledge assumptions about the behavior of players who are not fully sophisticated. This requires a novel approach. While in Kreps et al., Milgrom and Roberts, and the subsequent literature, the commitment of irrational types to a particular course of action renders high-order reasoning about behavior irrelevant, in our setting, players must entertain different conjectures about their opponent's behavior. This requires them to reason about their opponent's high-order beliefs. So, unlike in the existing literature, our results require an approach that is inherently strategic in nature. In particular, in our model, the assumption that there is a small amount of uncertainty about players' reasoning ability cannot be replaced by the assumption that there is a small amount of uncertainty about payoffs, unlike in the existing literature.

Finite depth of reasoning in Bayesian games. A growing literature studies the behavior of players with a finite depth of reasoning in games with incomplete information. The experimental literature studies behavior in a wide range of games ranging from auctions to betting and market entry games (see Crawford, Costa-Gomes, and Iriberri, 2013, for a survey). This

paper is the first to study the robustness of predictions to perturbations of beliefs about payoffs and reasoning ability. Strzalecki (2014) shows that the risk-dominant solution may not be uniquely selected in the electronic-mail game of Rubinstein (1989), but he does not consider the robustness of predictions. In fact, it is not possible to study the robustness of predictions in his framework since it does not allow for perturbations of beliefs about payoffs.[21] De Clippel, Saran, and Serrano (2014) study implementation problems using a variant of rationalizability. Their definition of rationalizability differs from ours in two important ways. First, De Clippel et al. assume that nonstrategic types follow a fixed strategy (truth-telling, in their setting) while we allow them to play any action, as in the level-k literature. Second, they assume that beliefs about payoffs and depth of reasoning are independent, while we allow for any correlation in beliefs. Kets (2010, 2013) introduces a class of type spaces for players with an arbitrary depth of reasoning and defines rationalizability and Bayesian-Nash equilibrium for her setting. An important difference is that Kets's solution concepts are refinements of the corresponding solution concepts for the standard case with common belief in an infinite depth.

Appendix A Type spaces

Belief hierarchies provide an explicit description of players' higher-order beliefs. Higher-order beliefs can also be described implicitly, using a type space (cf. Harsanyi, 1967). Here we define type spaces that generate belief hierarchies with an arbitrary (finite or infinite) depth. Type spaces are defined for a given set Θ of states of nature (taken to be compact metric, as before).

Definition 5. A *type space* is a tuple

$$\mathcal{T} := \left\langle \left(T_i\right)_{i=1,2}, \left(\beta_i\right)_{i=1,2}, \left(\chi_i\right)_{i=1,2} \right\rangle,$$

where for each player i, $T_i = T_i^\infty \cup \bigcup_{\ell=0}^\infty T_i^\ell$ is the set of *types* for player i, assumed to be nonempty and Polish, and χ_i is a continuous function that maps each type $t_i \in T_i$ into a *signal* $\chi_i(t_i) \in X_i$. The function β_i maps types into beliefs: for $t_i \in T_i^k$, $k \leq \infty$, $\beta_i(t_i) := \beta_i^k(t_i)$, where

- β_i^0 maps each $t_i \in T_i^0$ into $h_i^{*,0}$;
- β_i^k is continuous and maps each $t_i \in T_i^k$ into a belief in $\Delta(\Theta \times T_{-i}^{\leq k-1})$, where $T_{-i}^{\leq k} := \bigcup_{\ell=0}^k T_{-i}^\ell$;

[21] In addition, his universal type space does not contain the universal type space for standard types of Mertens and Zamir (1985), so that it is not possible to consider arbitrarily small deviations from standard assumptions in his framework.

- β_i^∞ is continuous and maps each $t_i \in T_i^\infty$ into a belief in $\Delta(\Theta \times T_{-i})$;

If there is $t_i \in T_i^k$ for $k = 1, 2, \ldots$, then there is a type $t_{-i} \in T_{-i}^m$ for some $m < k$.[22]

Thus, each type in T_i^0 is associated with the "naive" type $h_i^{*,0}$. Types in T_i^k, $k < \infty$, are mapped into a belief over the types with a strictly lower index, and types in T_i^∞ have a belief over types with any index. As before, a type's signal $\chi_i(t_i) \in X_i$ represents its (payoff-irrelevant) private information. A standard (Harsanyi) type space is simply a type space in which every type has index $k = \infty$. The online appendix shows that each type can be mapped into a belief hierarchy in the usual way, and that a type with index k corresponds to a belief hierarchy of depth k.

We can define the set of rationalizable actions for a type in the same way as before. The only difference is that the conjectures in the definition of the set $R_i^{T,k}$ of k-rationalizable actions are now (measurable) functions from $\Theta \times T_{-i}$ into $\Delta(A_{-i})$. As we show now, the set of (k-)rationalizable actions for a type depends only on its belief hierarchy, as in the standard case (Dekel, Fudenberg, and Morris, 2007, Lemma 1). To state the result, we need some more definitions. In the online appendix, we define a function h_i^T that maps each type into its (full) belief hierarchy. We then have:

Lemma A.1. *For any type $t_i \in T_i$, and every $k = 1, 2, \ldots$,*

$$R_i^{T,k}(t_i) = R_i^k(h_i^T(t_i)).$$

Hence,

$$R_i^{T,\infty}(t_i) = R_i^\infty(h_i^T(t_i)).$$

So, studying the rationalizable actions of players within the context of the universal type space is without loss of generality.

Appendix B The universal type space

Following Mertens and Zamir (1985), we use the set of all belief hierarchies to construct the universal type space. The set of types for player i is the set H_i of belief hierarchies. The next result implies that H_i is well-defined:

Lemma B.1. *For every $k = 1, 2, \ldots, \infty$, the set H_i^k is nonempty and compact metric. Hence, H_i is well-defined and Polish.*

[22] This ensures that types' beliefs are well-defined (i.e., have nonempty support).

The next result shows that each belief hierarchy specifies a belief about the full hierarchy of other players, not just about the individual levels of the hierarchy:

Lemma B.2.

(a) For each belief hierarchy $h_i = (x_i, \mu_i^0, \mu_i^1, \ldots) \in H_i^\infty$ there exists a unique Borel probability measure $\mu_i(h_i)$ on $\Theta \times H_{-i}$ such that

$$marg_{\widetilde{\Omega}_i^{\ell-1}} \mu_i(h_i) = \mu_i^\ell$$

for all $\ell = 1, 2, \ldots$.

(b) For each $k > 0$ and every belief hierarchy $h_i = (x_i, \mu_i^0, \mu_i^1, \ldots, \mu_i^k) \in H_i^k$, there exists a unique Borel probability measure $\mu_i(h_i)$ on $\Theta \times H_{-i}^{\leq k-1}$ such that

$$marg_{\widetilde{\Omega}_i^{\ell-1}} \mu_i(h_i) = \mu_i^\ell$$

for all $\ell = 1, \ldots, k$.

Thus, each belief hierarchy of player i can be associated with a belief over the set Θ of states of nature, the signal spaces X_{-i} and over the other players' belief hierarchies, in such a way that i's belief over her ℓth-order space of uncertainty coincides with his ℓth-order belief as specified by her hierarchy of beliefs. That is, the construction is *canonical*.

Using Lemma B.2, we can construct a function that assigns to each belief hierarchy h_i its signal (by projecting h_i onto X_i) and a belief about nature and other players' hierarchies (as given by Lemma B.2). The inverse of this function assigns to each signal-belief pair $(x_i, \mu_i) \in X_i \times \Delta(\Theta \times H_{-i})$ the associated belief hierarchy (possibly finite). Proposition B.3 shows that these functions are continuous. Thus, for every depth k, we have a homeomorphism.

Proposition B.3. *There is a homeomorphism $\tilde{\psi}_i^\infty : H_i^\infty \to X_i \times \Delta(\Theta \times H_{-i})$. Moreover, for each $k = 1, 2, \ldots$, there is a homeomorphism $\tilde{\psi}_i^k : H_i^k \to X_i \times \Delta\big(\Theta \times H_{-i}^{\leq k-1}\big)$.*

We use this result to define the universal type space. Write $\psi_i^k(h_i)$, $k = 1, 2, \ldots$, $h_i \in H_i^k$, for the projection of $\tilde{\psi}_i^k$ into $\Delta(\Theta \times H_{-i}^{\leq k-1})$; likewise, $\psi_i^\infty(h_i)$, $h_i \in H_i^\infty$, is the projection of $\tilde{\psi}_i^\infty(h_i)$ into $\Delta(\Theta \times H_{-i})$. Define $\psi_i^0 : H_i^0 \to \{h_i^{*,0}\}$ in the obvious way, and view $\psi_i^k(h_i)$, $h_i \in H_i^k$, $k < \infty$, as a probability measure on $\Delta(\Theta \times H_{-i})$. Let $\psi_i : H_i \to \Delta(\Theta \times H_{-i})$ be the function that coincides with ψ_i^k on H_i^k. Thus, ψ_i is continuous. Finally, let $\chi_i^*(x_i, \mu_i^0, \mu_i^1, \ldots) = x_i$. Then, $\mathcal{T}^* := \Big\langle \big(H_i\big)_{i=1,2}, \big(\psi_i\big)_{i=1,2}, (\chi_i^*)_{i=1,2} \Big\rangle$ is a type space. This type space is universal in the sense that it generates all belief hierarchies; see the online appendix for a proof. A useful property is that the universal type space is *complete* in the sense of Brandenburger (2003): for every belief $\nu_i \in \Delta(\Theta \times H_{-i})$, there is a type $h_i \in H_i$ with $\psi_{h_i} = \nu_i$.

Remark 2. Types with different depths of reasoning can look very similar, yet correspond to different beliefs. For example, consider a depth-k type h_i^k and a depth-m type h_i^m for $m < k$ that have the same beliefs about Θ and that both believe (i.e., assign probability 1 to the event) that the opponent has a given depth-$(m-1)$ type h_{-i}^{m-1}. These beliefs look very similar, yet they are different: the belief $\psi_{h_i^k}$ is a probability measure defined on the set $\bigcup_{\ell \leq k-1} H_{-i}^{\ell}$ of types of depth at most $k-1$, while the belief $\psi_{h_i^m}$ is a probability measure defined on the set $\bigcup_{\ell \leq m-1} H_{-i}^{\ell}$ of types of depth at most $m-1$. This distinction may seem merely technical, yet it has conceptual content: while the types h_i^k and h_i^m are behaviorally indistinguishable in a given game, they do correspond to different belief hierarchies. For example, when faced with an opponent with a high reasoning ability (i.e., with depth greater than $m-1$), type h_i^k could adapt its behavior, while type h_i^m would be unable to. (For an exploration of these issues in the context of the level-k literature, both theoretically and experimentally, see Agranov, Potamites, Schotter, and Tergiman (2012) and Alaoui and Penta (2016), amongst others.) $\triangleleft$

Appendix C Transfinite elimination process

We show that rationalizability satisfies a *best-reply property*: Any rationalizable action for a type is a best response against the belief that the opponent plays a rationalizable strategy, as in the standard case (Dekel, Fudenberg, and Morris, 2007, Prop. 4). We use this to show that the set of rationalizable actions cannot be refined further if we perform more rounds of elimination (Proposition C.2 below).

Lemma C.1. *For every $h_i \in H_i$ and $a_i \in R_i^{\infty}(h_i)$, there is a measurable conjecture $s_{-i} : \Theta \times H_{-i} \to \Delta(A_{-i})$ such that supp $s_{-i}(\theta, h_{-i}) \subset R_{-i}^{\infty}(h_{-i})$ for all θ, h_{-i}, and*

$$a_i \in \arg\max \max_{a_i' \in A_i} \int_{\Theta \times H_{-i} \times A_{-i}} u_i(a_i', a_{-i}, \theta) s_{-i}(\theta, h_{-i})(a_{-i}) \mathrm{d}\psi_{h_i}.$$

Proof. Fix $h_i \in H_i$ and $a_i \in R_i^{\infty}(h_i)$. The result follows directly if h_i has finite depth. So suppose that h_i has an infinite depth of reasoning. For every m, there exists $s_{-i}^m : \Theta \times H_{-i} \to \Delta(A_{-i})$ such that supp $s_{-i}^m(\theta, h_{-i}) \subset R_{-i}^{m-1}(h_{-i})$ for all θ, h_{-i}, and $a_i \in \operatorname{argmax}_{\tilde{a}_i} \int u_i(\tilde{a}_i, s_{-i}^m(\theta, h_{-i}), \theta) \mathrm{d}\psi_{h_i}$. We need to show that there is $s_{-i} : \Theta \times H_{-i} \to \Delta(A_{-i})$ such that supp $s_{-i}(\theta, h_{-i}) \subset R_{-i}^{\infty}(h_{-i})$ for all θ, h_{-i}, and $a_i \in \operatorname{argmax}_{\tilde{a}_i} \int u_i(\tilde{a}_i, s_{-i}(\theta, h_{-i}), \theta) \mathrm{d}\psi_{h_i}$. Since the set of functions from $\Theta \times H_{-i}$ to $\Delta(A_{-i})$ is compact Hausdorff, the sequence $\{s_{-i}^m\}_m$ has a convergent subsequence $\{s_{-i}^{m_k}\}_k$, and the (pointwise) limit $s_{-i} := \lim_{k \to \infty} s_{-i}^{m_k}$ is unique. Moreover, s_{-i} is measurable (Aliprantis and Border, 2005, Lemma 4.29). By construction, supp $s_{-i}(\theta, h_{-i}) \subset R_{-i}^{\infty}(h_{-i})$ for all θ, h_{-i}. By the dominated convergence theorem, $\int u_i(\tilde{a}_i, s_{-i}^{m_k}(\theta, h_{-i}), \theta) \mathrm{d}\psi_{h_i}$ converges to

$\int u_i(\tilde{a}_i, s_{-i}(\theta, h_{-i}), \theta)\mathrm{d}\psi_{h_i}$ for $\tilde{a}_i \in A_i$. Hence, $a_i \in \operatorname{argmax}_{\tilde{a}_i} u_i(\tilde{a}_i, s_{-i}^{m_k}(\theta, h_{-i}), \theta)\mathrm{d}\psi_{h_i}$. □

We can use this to show that continuing the elimination process does not eliminate any additional strategies. Following Lipman (1994), we define a rationalizability concept based on the transfinite elimination of strictly dominated strategies. This requires working with ordinals. Recall that an ordinal α can be identified with the set $\{\beta : \beta < \alpha\}$ of its predecessors; we identify the finite ordinals with the natural numbers $0, 1, 2, \ldots$, so that the first infinite ordinal ω is equal to $\{0, 1, \ldots\} = \mathbb{N}$. The successor of an ordinal α is the least ordinal greater than α. An ordinal is a *successor ordinal* if it is the successor of some ordinal. An ordinal is a *limit ordinal* if it is not 0 or a successor ordinal.

Let $\mathcal{R}_i^0 = R_i^0$. For $\alpha > 0$, suppose $\mathcal{R}_i^\gamma$ has been defined for all $\gamma < \alpha$. If α is the successor of some ordinal β, then the definition of $\mathcal{R}_i^\alpha$ is similar to before:

$$\mathcal{R}_i^\alpha(h_i) := \left\{ a_i \in A_i : \begin{array}{l} \text{there is a measurable } s_{-i} : \Theta \times H_{-i} \to \Delta(A_{-i}) \text{ s.t.} \\ \text{supp } s_{-i}(\theta, h_{-i}) \subseteq \mathcal{R}_{-i}^\beta(h_{-i}) \text{ for all } h_{-i} \in H_{-i}, \theta \in \Theta; \text{ and} \\ a_i \in \arg\max_{a_i' \in A_i} \int_{\Theta \times H_{-i} \times A_{-i}} u_i(a_i', a_{-i}, \theta) s_{-i}(\theta, h_{-i})(a_{-i}) \mathrm{d}\psi_{h_i} \end{array} \right\}.$$

If α is a limit ordinal, then define $\mathcal{R}_i^\alpha$ by

$$\mathcal{R}_i^\alpha(h_i) := \bigcap_{\gamma < \alpha} \mathcal{R}_i^\gamma(h_i).$$

So, for finite α, $\mathcal{R}_i^\alpha(h_i) = R_i^\alpha(h_i)$. Moreover, $\mathcal{R}_i^\omega(h_i) = R_i^\infty(h_i)$. As we iterate beyond ω, the set of rationalizable actions may continue to shrink. It turns out, though, that performing transfinitely many rounds of elimination of strictly dominated strategies does not eliminate any additional actions:

Proposition C.2. *For any ordinal $\alpha \geq \omega$ and any $h_i \in H_i$, $\mathcal{R}_i^\alpha(h_i) = R_i^\infty(h_i)$.*

Proof. By Lemma C.1, $\mathcal{R}_i^{\omega+1}(h_i) = R_i^\infty(h_i)$. □

Appendix D Proofs

D.1 Proof of Lemma 3.3

The result follows from a simple induction. For $h_i \in H_i^0$, $R_i^0(h_i) = R_i^1(h_i) = \cdots = R_i^\infty(h_i)$. For $m > 0$, suppose that for $n \leq m-1$, $h_i \in H_i^n$, $R_i^n(h_i) = R_i^{n+1}(h_i) = \cdots = R_i^\infty(h_i)$. Then, for any $h_i \in H_i^m$, $R_i^m(h_i) = R_i^{m+1}(h_i) = \cdots = R_i^\infty(h_i)$, as ψ_{h_i} has support in $H_{-i}^0 \cup \cdots \cup H_{-i}^{m-1}$. □

D.2 Proof of Lemma 3.4

The proof of Lemma 3.4 below is a straightforward adaptation of the proof of Lemma 3 in Yildiz (2005). Lemma 3 of Yildiz (2005) extends Lemma 1 of Dekel, Fudenberg, and Morris (2007) to the case where Θ is compact metric (rather than finite, as in Dekel et al.). The only significant difference between Yildiz's setting and ours is that the set H_i of belief hierarchies of arbitrary depth satisfies weaker topological conditions than the set of standard belief hierarchies that Yildiz considers: the set H_i is Polish, while the set of standard belief hierarchies is compact metric. We therefore adapt Yildiz's proof so that it makes reference only to the set of finite-order belief hierarchies, which satisfy the same topological conditions as in the standard case.

We define a version of m-rationalizability that is a function only of players' mth-order belief hierarchies, and then use it to prove the results for R_i^m (whose domain is the set H_i of belief hierarchies). We need some more notation. For $m = 0, 1, \ldots$, define $G_i^m := H_i^m \cup \widetilde{H}_i^m$ to be the set of mth-order belief hierarchies. Also, define $\tilde{G}_i^m := H_i^0 \cup \cdots \cup H_i^{m-1} \cup G_i^m$. For an mth-order belief hierarchy $g_i^m = (x_i, \mu_i^0, \ldots, \mu_i^m)$, write $\nu_{g_i^m}^m := \mu_i^m$ for its induced mth-order belief. Note that $\nu_{g_i^m}^m$ is a belief on $\Theta \times \tilde{G}_i^{m-1}$. For $g_i \in \tilde{G}_i^m$, define $n_{g_i} = k$ if $g_i \in H_i^k$ and $n_{g_i} = m$ if $g_i \in \widetilde{H}_i^m$.

For $g_i^0 \in G_i^0$, let $\widetilde{R}_i^0(g_i^0) := A_i$. For $m > 0$, suppose that $\widetilde{R}_i^{m-1} : G_i^{m-1} \twoheadrightarrow A_i$ has been defined and define the correspondence $\widetilde{R}_i^m : G_i^m \twoheadrightarrow A_i$ by

$$\widetilde{R}_i^m(g_i^m) := \left\{ a_i \in A_i : \begin{array}{l} \text{there is a measurable } s_{-i} : \Theta \times \tilde{G}_{-i}^{m-1} \to \Delta(A_{-i}) \text{ s.t.} \\ \text{supp } s_{-i}(\theta, g_{-i}) \subseteq \widetilde{R}_{-i}^{n_{g_i}}(g_{-i}) \text{ for all } g_{-i} \in \tilde{G}_{-i}^{m-1}, \theta \in \Theta; \text{ and} \\ a_i \in \arg\max_{a_i' \in A_i} \int_{\Theta \times \tilde{G}_{-i}^{m-1} \times A_{-i}} u_i(a_i', a_{-i}, \theta) s_{-i}(\theta, g_{-i})(a_{-i}) \mathrm{d}\nu_{g_i^m}^m. \end{array} \right\}.$$

We show that $\widetilde{R}_i^m$ is upper hemicontinuous,[23] We then use this to prove that R_i^m is upper hemicontinuous for $m \leq \infty$.

Lemma D.1. *The correspondence $\widetilde{R}_i^m$ is upper hemicontinuous and has nonempty values.*

Proof. By Theorem 17.11 of Aliprantis and Border (2005), $\widetilde{R}_i^m$ is upper hemicontinuous if and only if it has a closed graph, where the graph of a correspondence $F : X \twoheadrightarrow Y$ is $\mathrm{Gr}F = \{(x, y) \in X \times Y : y \in F(x)\}$. By definition, $\mathrm{Gr}(\widetilde{R}_i^0) = G_i^0 \times A_i$. So, $\widetilde{R}_i^0$ is nonempty-valued and upper hemicontinuous. For $m > 0$, suppose that $\widetilde{R}_i^n$ is upper hemicontinuous and nonempty-valued for $n \leq m - 1$. We claim that $\mathrm{Gr}(\widetilde{R}_i^m)$ is closed. By the induction hypothesis, $\Theta \times \bigcup_{n \leq m-1} \mathrm{Gr}(\widetilde{R}_{-i}^n) \subset \Theta \times \tilde{G}_{-i}^{m-1} \times A_{-i}$ is closed and nonempty. Since $\Theta \times \tilde{G}_{-i}^{m-1} \times A_{-i}$ is compact,

[23] Recall that a correspondence $F : X \twoheadrightarrow Y$ is upper hemicontinuous if and only if $\{x \in X : F(x) \subset U\}$ for every open subset of Y.

so is $\Theta \times \bigcup_{n \leq m-1} \mathrm{Gr}(\widetilde{R}^n_{-i})$. Hence, $\Delta(\Theta \times \bigcup_{n \leq m-1} \mathrm{Gr}(\widetilde{R}^n_{-i}))$ is compact. Moreover, since u_i is continuous and bounded (being defined on a compact space), $\int u_i(\cdot, a_{-i}, \theta) d\nu^m_i$ is a continuous function of $\nu^m_i \in \Delta(\Theta \times \bigcup_{n \leq m-1} \mathrm{Gr}(\widetilde{R}^n_{-i}))$ to $\mathbb{R}$. Define $\mathcal{M}^m_i : \Delta(\Theta \times \bigcup_{n \leq m-1} \mathrm{Gr}(\widetilde{R}^n_{-i})) \twoheadrightarrow A_i$ by

$$\mathcal{M}^m_i(\nu^m_i) := \mathrm{argmax}_{\tilde{a}_i \in A_i} \int u_i(\tilde{a}_i, a_{-i}, \theta) \mathrm{d}\nu^m_i.$$

By the Berge maximum theorem, $\mathcal{M}^m_i(\nu^m_i)$ is nonempty, and $\mathrm{Gr}(\mathcal{M}^m_i)$ is closed and thus compact in $\Delta(\Theta \times \bigcup_{n \leq m-1} \mathrm{Gr}(\widetilde{R}^n_{-i})) \times A_i$. Fix $\nu^m_i \in \Delta(\Theta \times \bigcup_{n \leq m-1} \mathrm{Gr}(\widetilde{R}^n_{-i}))$. If ν^m_i has support in $\Theta \times (H^0_{-i} \cup \cdots \cup H^{m-1}_{-i}) \subsetneq \Theta \times \tilde{G}^{m-1}_{-i}$, then there exist unique $g^m_i \in H^m_i$ and $\tilde{g}^m_i \in \widetilde{H}^m_i$ such that the marginal of ν^m_i on Θ and the other player's $(m-1)$th-order belief hierarchy coincides with the mth-order belief induced by g^m_i and $\tilde{g}^m_i$ (Remark 2). Otherwise (if ν^m_i has support in $\Theta \times (H^0_{-i} \cup \cdots \cup H^{m-1}_{-i} \cup \widetilde{H}^{m-1}_{-i}) = \Theta \times \tilde{G}^{m-1}_{-i}$), $\tilde{g}^m_i$ is the unique mth-order belief hierarchy in G^m_i such that the induced mth-order belief coincides with the marginal of ν^m_i. Let φ^m_i and $\tilde{\varphi}^m_i$ be the functions that map (a_i, ν^m_i) into (a_i, g^m_i) and $(a_i, \tilde{g}^m_i)$, respectively (if the former exist). Then, φ^m_i and $\tilde{\varphi}^m_i$ are continuous (on the appropriate domain and range spaces), and $\mathrm{Gr}(\widetilde{R}^m_i) = \varphi_i(\mathcal{M}^m_i) \cup \tilde{\varphi}^m_i(\mathcal{M}^m_i)$. So, $\mathrm{Gr}(\widetilde{R}^m_i)$ is closed and $\widetilde{R}^m_i$ is upper hemicontinuous (Aliprantis and Border, 2005, Thm. 17.23, 17.24). It then follows from standard extension theorems (e.g., Lubin, 1974) that $\widetilde{R}^m_i(g^m_i)$ is nonempty for any $g^m_i \in G^m_i$. □

We next relate R^m_i to $\widetilde{R}^m_i$ and show that R^m_i is upper hemicontinuous. Write $\mathrm{proj}_{G^m_i}$ for the projection function from the set $H^\infty_i \cup \bigcup_{k \geq m} H^k_i$ of belief hierarchies of depth at least m into G^m_i. For any $h_i \in H^\infty_i \cup \bigcup_{k \geq m} H^k_i$, define $\tilde{h}^m_i := \mathrm{proj}_{G^m_i}(h_i)$. Then:

Lemma D.2. *For any $h_i \in H_i$,*

$$R^m_i(h_i) = \begin{cases} \widetilde{R}^m_i(\tilde{h}^m_i) & \text{if } h_i \in H^k_i \text{ for } k \geq m; \\ \widetilde{R}^k_i(\tilde{h}^k_i) & \text{if } h_i \in H^k_i \text{ for } k < m. \end{cases}$$

The proof is similar to that of Lemma A.1 and thus omitted. It is now immediate that R^m_i is nonempty and upper hemicontinuous: Fix $A'_i \subset A_i$. Then,

$$\{h_i \in H_i : R^m_i(h_i) \subset A'_i\} = \bigcup_{k=m,m+1,\ldots,\infty} \{h_i \in H^k_i : \widetilde{R}^m_i(\tilde{h}^m_i) \subset A'_i\} \cup \bigcup_{k \leq m-1} \{h_i \in H^k_i : \widetilde{R}^m_i(\tilde{h}^k_i) \subset A'_i\}.$$

Each of the sets $\{h_i \in H^k_i : \widetilde{R}^m_i(\tilde{h}^k_i) \subset A'_i\}$ is open (Aliprantis and Border, 2005, Thm. 17.23). So, $\{h_i \in H_i : R^m_i(h_i) \subset A'_i\}$ is open, and R^m_i is upper hemicontinuous. Consequently, R^∞_i is upper hemicontinuous (Aliprantis and Border, 2005, Thm. 17.25). □

D.3 Proof of Lemma 4.2

Since ψ_i and the function that maps ψ_{h_i} into its marginal on Θ are both continuous (Proposition 3.1 and Aliprantis and Border 2005, Thm. 15.14), it suffices to show that the set $\Delta_i^{A'}$ is nonempty and open in $\Delta(\Theta)$. By assumption, $\Delta_i^{A'}$ is nonempty. So, it remains to show that every element of $\Delta_i^{A'}$ has a neighborhood in $\Delta_i^{A'}$.

The first step is to dispose of the quantification in 1 and 2 in Assumption R-Mult(A'). By Berge's maximum theorem (Aliprantis and Border, 2005, Thm. 17.31), for every action $a_i \in A_i$ for i and for every nonempty subset $B_{-i} \subset A_{-i}$ of actions for the opponent, the correspondence that maps $\theta \in \Theta$ to $\text{argmax}\{u_i(a_i, a_{-i}, \theta) : a_{-i} \in B_{-i}\}$ is upper hemicontinuous. By the Kuratowski-Ryll-Nardzewski selection theorem (Aliprantis and Border, 2005, Thm. 18.13), this correspondence admits a measurable selection. So, for every $a_i \in A_i$ and $B_{-i} \subset A_{-i}$, we can fix a measurable function $q_{-i}(\cdot \mid a_i, B_{-i}) : \Theta \to B_{-i}$ such that for all $\theta \in \Theta$, $u_i(a_i, q_{-i}(\theta \mid a_i, B_{-i}), \theta) \geq u_i(a_i, a_{-i}, \theta)$ for all $a_{-i} \in B_{-i}$. So, $u_i(a_i, q_{-i}(\cdot \mid a_i, B_{-i}), \cdot)$ is Borel measurable. We can think of $q_{-i}(\cdot \mid a_i, B_{-i})$ as the "conjecture" about the opponent's beliefs that gives the highest payoff for a_i for every state $\theta \in \Theta$ given that the opponent chooses an action in B_{-i}. If $a_i \in A'_i$, the relevant case is the one where the opponent chooses an action in A'_{-i}; if $a_i \notin A'_i$, we want to allow the opponent to play any action $a_{-i} \in A_{-i}$ (cf. Assumption R-Mult(A')).

We can now rewrite the conditions in Assumption R-Mult(A') without quantifying over measurable functions $\widetilde{s}_{-i}$. Fix $\mu_i \in \Delta_i^{A'}$. Then,

1. for all $a'_i \in A'_i$,
$$\{a'_i\} = \text{argmax}_{\tilde{a}_i \in A_i} \int_\Theta u_i(\tilde{a}_i, q_{-i}(\theta \mid a'_i, A'_{-i}), \theta) d\mu_i;$$
2. for all $a''_i \notin A'_i$,
$$a''_i \notin \text{argmax}_{\tilde{a}_i \in A_i} \int_\Theta u_i(\tilde{a}_i, q_{-i}(\theta \mid a''_i, A_{-i}), \theta) d\mu_i$$

We can now bound the payoff differences. For $a'_i \in A'_i$, define

$$\xi_{a'_i} := \min_{a_i \neq a'_i} \Big\{ \int_\Theta u_i(a'_i, q_{-i}(\theta \mid a'_i, A'_{-i}), \theta) d\mu_i - \int_\Theta u_i(a_i, q_{-i}(\theta \mid a'_i, A'_{-i}), \theta) d\mu_i \Big\},$$

so $\xi_{a'_i} > 0$. For $a''_i \notin A'_i$, define

$$\zeta_{a''_i} := \min_{a_i \neq a''_i} \Big\{ \max\Big\{0, \int_\Theta u_i(a_i, q_{-i}(\theta \mid a''_i, A_{-i}), \theta) d\mu_i - \int_\Theta u_i(a''_i, q_{-i}(\theta \mid a''_i, A_{-i}), \theta) d\mu_i \Big\}\Big\}.$$

Again, $\zeta_{a''_i} > 0$. Note that since u_i is a continuous function on a compact space, there is $c > 0$ such that $u_i(a_i, a_{-i}, \theta) \in [-\frac{c}{2}, \frac{c}{2}]$ for all a_i, a_{-i}, θ.

The next step is to construct a neighborhood of μ_i. We cannot do so directly using the integral of $u_i(\cdot, q_{-i}(\cdot \mid a_i, B_{-i}), \cdot)$ (for $a_i \in A_i$ and $B_{-i} \subset A_{-i}$) because this function may not

be continuous if Θ is uncountably infinite. We can, however, approximate it by a continuous function. By Lusin's theorem (Aliprantis and Border, 2005, Thm. 12.8), for each $\eta > 0$, there is a compact subset $K_\eta \subset \Theta$ such that $\mu_i(K_\eta) > 1 - \eta$ and for all $a_i, a_i' \in A_i$, $B_{-i} \subset A_{-i}$, the restriction of $u_i(a_i, q_{-i}(\cdot \mid a_i', B_{-i}), \cdot)$ to K_η, denoted $u_i^\eta(a_i, q_{-i}(\cdot \mid a_i', B_{-i}), \cdot)$, is continuous. By Tietze's extension theorem (Aliprantis and Border, 2005, Thm. 2.47), each function $u_i^\eta(a_i, q_{-i}(\cdot \mid a_i', B_{-i}), \cdot)$ has a continuous extension $\tilde{u}_i^\eta(a_i, q_{-i}(\cdot \mid a_i', B_{-i}), \cdot)$ to Θ.

We are now ready to define a neighborhood of μ_i in $\Delta(\Theta)$. For $\tilde{a}_i \in A_i$, let $\delta^{\tilde{a}_i} := (\delta_{a_i}^{\tilde{a}_i})_{a_i \in A_i}$ be such that $\delta_{a_i}^{\tilde{a}_i} > 0$ for all $a_i \in A_i$, and let $\delta := (\delta^{\tilde{a}_i})_{\tilde{a}_i \in A_i}$. For $a_i' \in A_i'$, define

$$O_i(\mu_i; a_i', \eta, \delta) := \bigcap_{a_i \in A_i} \Big\{ \nu_i \in \Delta(\theta) : \Big| \int_\Theta \tilde{u}_i^\eta(a_i, q_{-i}(\theta \mid a_i', A_{-i}'), \theta) d\nu_i - \int_\Theta \tilde{u}_i^\eta(a_i, q_{-i}(\theta \mid a_i', A_{-i}'), \theta) d\mu_i \Big| < \delta_{a_i}^{a_i'} \Big\}.$$

Then, $O_i(\mu_i; a_i', \eta, \delta)$ is open (Billingsley, 1968, App. III). Moreover, it contains μ_i. For $a_i'' \notin A_i'$, define

$$O_i(\mu_i; a_i'', \eta, \delta) := \bigcap_{a_i \in A_i} \Big\{ \nu_i \in \Delta(\Theta) : \Big| \int_\Theta \tilde{u}_i^\eta(a_i, q_{-i}(\theta \mid a_i'', A_{-i}), \theta) d\nu_i - \int_\Theta \tilde{u}_i^\eta(a_i, q_{-i}(\theta \mid a_i'', A_{-i}), \theta) d\mu_i \Big| < \delta_{a_i}^{a_i''} \Big\}.$$

Again, $O_i(\mu_i; a_i'', \eta, \delta)$ is open and it contains μ_i. Define

$$O_i(\mu_i; \eta, \delta) := \bigcap_{a_i \in A_i} O_i(\mu_i; a_i, \eta, \delta).$$

Then, $O_i(\mu_i; \eta, \delta)$ is open and it contains μ_i. We can then choose $\hat{\eta} > 0$ and $\hat{\delta} > 0$ such that if $\nu_i \in O_i(\mu_i; \hat{\eta}, \hat{\delta})$, then (i) for all $a_i' \in A_i'$, $a_i \neq a_i'$,

$$\int_\Theta u_i(a_i', q_{-i}(\theta \mid a_i', A_{-i}'), \theta) \mathrm{d}\nu_i - \int_\Theta u_i(a_i, q_{-i}(\theta \mid a_i', A_{-i}'), \theta) \mathrm{d}\nu_i \geq \xi_{a_i'} - c \cdot \hat{\eta} - \hat{\delta}_{a_i}^{a_i'} - \hat{\delta}_{a_i'}^{a_i'} > 0;$$

and (ii) for all $a_i'' \notin A_i'$, $a_i \neq a_i''$,

$$\int_\Theta u_i(a_i, q_{-i}(\theta \mid a_i'', A_{-i}), \theta) \mathrm{d}\nu_i - \int_\Theta u_i(a_i'', q_{-i}(\theta \mid a_i'', A_{-i}), \theta) \mathrm{d}\nu_i \geq \zeta_{a_i''} - c \cdot \hat{\eta} - \hat{\delta}_{a_i}^{a_i''} - \hat{\delta}_{a_i''}^{a_i''} > 0.$$

Then, $O_i(\mu_i; \hat{\eta}, \hat{\delta}) \subset \Delta_i^{A'}$. So, $\Delta_i^{A'}$ is open. $\square$

D.4 Proof of Lemma 4.3

Let $m = 0, 1, 2, \ldots, \infty$. Suppose that $V_{-i} \subset H_{-i}$ is open and that for all $h_{-i} \in V_{-i}$, $R^m_{-i}(h_{-i}) = A'_{-i}$. Let $h_i \in H_i$ be a type with $\text{marg}_\Theta \psi_{h_i} \in \Delta_i^{A'}$ and $\psi_{h_i}(V_{-i}) = 1$. Then, clearly, $R_i^{m+1}(h_i) = A'_i$: h_i assigns probability 1 to types that can play precisely the actions in A'_{-i}, and the actions that are a best response are precisely the actions in A'_i. It remains to show that h_i has a neighborhood such that $R_i^{m+1}(h'_i) = A'_i$ for every type h'_i in the neighborhood. It will be convenient to define

$$D_i(a_i, a'_i, a_{-i}, \theta) := u_i(a_i, a_{-i}, \theta) - u_i(a'_i, a_{-i}, \theta)$$

for $a_i, a'_i \in A_i$, $a_{-i} \in A_{-i}$, and $\theta \in \Theta$. The definition can be extended to mixed strategies in the obvious way. Since the (m-)rationalizability correspondence is upper hemicontinuous (Lemma 3.4), it follows from the Kuratowski-Ryll-Nardzewski selection theorem (Aliprantis and Border, 2005, Thm. 18.13) that for each $a'_i \in A'_i$, there is a (measurable) conjecture $s_{-i}^{a'_i} : \Theta \times H_{-i} \to \Delta(A_{-i})$ such that

$$\begin{aligned}
&\text{supp}\, s_{-i}^{a'_i}(\theta, h_{-i}) \subset R^{m-1}_{-i}(h_{-i}) \text{ for all } \theta, h_{-i};\\
&\int D_i(a'_i, a_i, s_{-i}^{a'_i}(\theta, h), \theta) \mathrm{d}\psi_{h_i} > 0 \text{ for all } a_i \neq a'_i;\\
&s_{-i}^{a'_i}(\theta, h_{-i}) = q_{-i}(\theta \mid a'_i, A'_{-i}) \text{ for } h_{-i} \in V^{m-1}_{-i};
\end{aligned}$$

where $q_{-i}(\theta \mid a'_i, A'_{-i})$ has been defined in the proof of Lemma 4.2. Likewise, for $a''_i \notin A'_i$, for every measurable conjecture $s_{-i}^{a''_i} : \Theta \times H_{-i} \to \Delta(A_{-i})$ such that $\text{supp}\, s_{-i}^{a''_i}(\theta, h_{-i}) \subset R^{m-1}_{-i}(h_{-i})$, there is $a_i \neq a''_i$ such that

$$\int D_i(a''_i, a_i, s_{-i}^{a''_i}(\theta, h), \theta) \mathrm{d}\psi_{h_i} < 0.$$

For $\upsilon > 0$, define

$$O_i(h_i; \upsilon) := \{h'_i \in H_i : \psi_{h'_i}(V_{-i}) > \psi_{h_i}(V_{-i}) - \upsilon\}.$$

This set is open (Billingsley, 1968, App. III), and it contains h_i. For $\eta > 0$ and $\delta = (\delta_{a_i}^{\tilde{a}_i})_{a_i, \tilde{a}_i \in A_i}$ with $\delta_{a_i}^{\tilde{a}_i} > 0$ for $a_i, \tilde{a}_i$, define

$$O_i(h_i; \eta, \delta) := \{h'_i \in H_i : \text{marg}_\Theta \psi_{h'_i} \in O_i(\text{marg}_\Theta \psi_{h_i}; \eta, \delta)\}$$

where we have again used the notation from Lemma 4.2. Again, this set is open and contains h_i. So, the set

$$O_i(h_i; \eta, \delta, \upsilon) := O_i(h_i; \upsilon) \cap O_i(h_i; \eta, \delta)$$

is nonempty and open. By a similar argument as in the proof of Lemma 4.2, if we choose $\hat{\eta}, \hat{\upsilon}, \hat{\delta} > 0$ sufficiently close to 0, the actions in A'_i are (strictly) m-rationalizable for the types in $O_i(h_i; \hat{\eta}, \hat{\delta}, \hat{\upsilon})$, and no other actions are m-rationalizable for these types (i.e., $R_i^m(h'_i) = A'_i$ for $h'_i \in O_i(h_i; \hat{\eta}, \hat{\delta}, \hat{\upsilon})$). Let $O(h_i) := O_i(h_i; \hat{\eta}, \hat{\delta}, \hat{\upsilon})$. □

D.5 Proof of Lemma 4.4

By Lemma 4.2, the set $S_i^1 := \{h_i \in H_i : R_i^1(h_i) = A_i'\}$ is a nonempty and open. In particular, the interior $U_i^1 := S_i^1$ of S_i^1 is nonempty. So, A_i' is robustly 1-rationalizable for the types in U_i^1. Since $\mathcal{T}^*$ is complete (Appendix B), there is a type $h_i \in U_i^1$ that assigns probability 1 to U_{-i}^1.

For $m > 1$, suppose that the set $S_i^{m-1} := \{h_i \in H_i : R_i^{m-1}(h_i) = A_i'\}$ has a nonempty open subset U_i^{m-1} and that there is a type $h_i \in U_i^{m-1}$ that assigns probability 1 to U_{-i}^{m-1}. By construction, h_i has a belief in the multiplicity set (i.e., $h_i \in \Delta_i^{A'}$). By Lemma 4.3, h_i has a neighborhood $O(h_i)$ such that $R_i^m(h_i') = A_i'$ for $h_i' \in O(h_i)$. So, A_i' is robustly m-rationalizable for h_i. Let U_i^m be the union of such neighborhoods $O(h_i)$ (where h_i ranges over the types in U_i^{m-1} that assign probability 1 to U_{-i}^{m-1}). So, U_i^m is a nonempty open subset of $S_i^m := \{h_i \in H_i : R_i^\infty(h_i) = A_i'\}$. Again, by completeness, there exist types in U_i^m that assign probability 1 to U_{-i}^m. □

D.6 Proof of Theorem 4.1 (ctnd)

Consider a type $h_i^0 \in B_i^0$ with beliefs in the multiplicity set $\Delta_i^{A'}$ that assigns probability 1 to $\bigcup_m V_{-i}^m$. (Such a type exists since $\mathcal{T}^*$ is complete; see Appendix B.) By Lemma 4.3, h_i^0 has a neighborhood $O(h_i^0)$ such that $R_i^\infty(h_i') = A_i'$ for $h_i' \in O(h_i^0)$. So, A_i' is robustly rationalizable for h_i^0. For $n > 0$, suppose that there is a type $h_{-i}^{n-1} \in B_{-i}^{n-1}$ such that A_{-i}' is robustly rationalizable for h_{-i}^{n-1}. That is, h_{-i}^{n-1} has a neighborhood $O(h_{-i}^{n-1})$ such that $R_{-i}^\infty(h_{-i}) = A_{-i}'$. Consider a type $h_i^n \in B_i^n$ that assigns probability 1 to $O(h_{-i}^{n-1})$. Then, by Lemma 4.3, A_i' is robustly rationalizable for h_i. □

D.7 Proof of Lemma 4.6

We use the following auxiliary result:

Claim D.3. *There exist $\underline{x}_1^\varepsilon, \underline{x}_2^\varepsilon, \ldots$ and $\underline{x}_1^\varepsilon, \underline{x}_2^\varepsilon, \ldots$, such that for every k, $0 \le \underline{x}_k^\varepsilon < \frac{1}{2} < \bar{x}_k^\varepsilon \le 1$ and for every depth-k type h_i in T^ε with signal $x_i \in X_i^\varepsilon$,*

- *not investing is the unique rationalizable action whenever $x_i < \underline{x}_k^\varepsilon$;*
- *investing is the unique rationalizable actions whenever $x_i > \bar{x}_k^\varepsilon$;*
- *both actions are rationalizable whenever $x_i \in [\underline{x}_k^\varepsilon, \bar{x}_k^\varepsilon]$.*

Moreover, for every k, $\underline{x}_k^\varepsilon \to 0$ and $\bar{x}_k^\varepsilon \to 1$ as $\varepsilon \to 0$.

Proof of claim. Consider a depth-1 type h_i in T^ε. Then, NI is the unique rationalizable action for h_i if and only if its signal x_i is less than 0; and I is its unique rationalizable action if and only if $x_i > 1$. So, let $\underline{x}_1^\varepsilon := 0$ and $\bar{x}_1^\varepsilon := 1$. For $k > 1$, suppose the claim is true for $k-1$. If the game has complete information (i.e., $\varepsilon = 0$), then we can set $\underline{x}_k = \underline{x}_{k-1}$ and $\bar{x}_k = \bar{x}_{k-1}$, and we are done. So suppose $\varepsilon > 0$. Consider a type in T^ε with signal $x_i \in [-1+\varepsilon, 2-\varepsilon]$, and fix $z \in [x_i - \varepsilon, x_i]$. The posterior probability that the type assigns to the opponent having signal $x_{-i} \leq z$ is then

$$\pi^\varepsilon(z; x_i) = \int_{x_i-\varepsilon}^{z+\varepsilon} \left(\int_{\theta-\varepsilon}^{z} \frac{dx_{-i}}{2\varepsilon} \right) \frac{d\theta}{2\varepsilon} = \frac{1}{8\varepsilon^2}(z - x_i + 2\varepsilon)^2.$$

If the type has depth k, its expected payoff to I is at most

$$(1 - \pi^\varepsilon(\underline{x}_{k-1}^\varepsilon; x_i)) \cdot x_i + \pi^\varepsilon(\underline{x}_{k-1}^\varepsilon; x_i) \cdot (x_i - 1);$$

and it is at least

$$(1 - \pi^\varepsilon(\bar{x}_{k-1}^\varepsilon; x_i)) \cdot x_i + \pi^\varepsilon(\bar{x}_{k-1}^\varepsilon; x_i) \cdot (x_i - 1).$$

So, not investing is the unique rationalizable action for the type if $x_i < \underline{x}_k^\varepsilon$ and investing is the unique rationalizable action if $x_i > \bar{x}_k^\varepsilon$, where $x_k^\varepsilon = \underline{x}_k^\varepsilon, \bar{x}_k^\varepsilon$ solves $x_k^\varepsilon = \pi^\varepsilon(x_{k-1}^\varepsilon; x_k^\varepsilon)$, that is,

$$x_k^\varepsilon = 4\varepsilon^2 + 2\varepsilon + x_{k-1}^\varepsilon - 4\varepsilon\sqrt{\varepsilon^2 + \tfrac{1}{2}x_{k-1}^\varepsilon + \varepsilon}.$$

The function $f^\varepsilon(z) = 4\varepsilon^2 + 2\varepsilon + z - 4\varepsilon\sqrt{\varepsilon^2 + \frac{1}{2}z + \varepsilon}$ is increasing in z, with $f^\varepsilon(0) > 0$ and $f^\varepsilon(1) < 1$. Finally, the equation $f^\varepsilon(z) = z$ has a unique solution at $z = \frac{1}{2}$. This proves the first part of the claim.

We next show that the thresholds $\underline{x}_k^\varepsilon$ and $\bar{x}_k^\varepsilon$ converge to 0 and 1, respectively, as $\varepsilon \to 0$. To show this, note that $\lim_{\varepsilon\to 0} f^\varepsilon(\underline{x}_1^\varepsilon) = 0$ and $\lim_{\varepsilon\to 0} f^\varepsilon(\bar{x}_1^\varepsilon) = 1$. For $k > 1$, suppose, inductively, that $\lim_{\varepsilon\to 0} f^\varepsilon(\underline{x}_{k-1}^\varepsilon) = 0$ and $\lim_{\varepsilon\to 0} f^\varepsilon(\bar{x}_{k-1}^\varepsilon) = 1$. Then, it follows directly that $\lim_{\varepsilon\to 0} f^\varepsilon(\underline{x}_k^\varepsilon) = 0$ and $\lim_{\varepsilon\to 0} f^\varepsilon(\bar{x}_k^\varepsilon) = 1$. □

By Claim D.3, both actions are strictly rationalizable for a depth-k type in T^ε whenever its signal x_i lies strictly between $\underline{x}_k^\varepsilon$ and $\bar{x}_k^\varepsilon$. It remains to show that multiplicity is robust. Consider a depth-1 type h_i in T^ε with $x_i \in (\underline{x}_1^\varepsilon, \bar{x}_1^\varepsilon)$. For $\delta > 0$, define

$$O_i(x_i, 1; \delta) := \{h_i \in H_i^1 : \int \theta \mathrm{d}\psi_{h_i} \in (x_i - \delta, x_i + \delta)\}.$$

Then, $O_i(x_i, 1; \delta)$ contains h_i and is open in H_i (Billingsley, 1968, App. III). If we choose $\delta^* > 0$ sufficiently small, then both actions are strictly rationalizable for the types in $O_i(x_i, 1; \delta^*)$. Define $O_{x_i,1} := O_i(x_i, 1; \delta^*)$. So, multiplicity is robust for h_i.

For $k > 1$, the proof is a little more involved because depth-k types have nontrivial beliefs about the rationalizable strategies of the opponent. Fix a depth-k type h_i in T^ε with signal $x_i \in (\underline{x}^\varepsilon_k, \bar{x}^\varepsilon_k)$ and write $p_{x_i,k}$ for the probability that h_i assigns to the opponent having a signal in $(\underline{x}^\varepsilon_{k-1}, \bar{x}^\varepsilon_{k-1})$. Since both actions are strictly rationalizable for h_i, we have $p_{x_i,k} > x_i, 1 - x_i$. Since ψ_{h_i} is regular, for every $\eta > 0$, there is a compact subset K_η of the set of types in $T^{\varepsilon,k-1}_{-i}$ with signal $x_{-i} \in (\underline{x}^\varepsilon_{k-1}, \bar{x}^\varepsilon_{k-1})$ such that $\psi_{h_i}(K_\eta) > p_{x_i,k} - \eta$. Since K_η is compact, it has a (finite) open cover $V_{x_i,k,\eta} := \bigcup_{m=1}^{\ell} O_{x_i^m,k-1}$, where $x_i^m \in (\underline{x}^\varepsilon_{k-1}, \bar{x}^\varepsilon_{k-1})$. So, $\psi_{h_i}(V_{x_i,k,\eta}) > p_{x_i,k} - \eta$. For $\xi, \delta > 0$, define

$$O_i(x_i, k; \eta, \xi, \delta) := \{h'_i \in H_i : \psi_{h'_i}(V_{x_i,k,\eta}) > p_{x_i,k} - \eta - \xi\} \cap \\ \{h'_i \in H_i : \int \theta \mathrm{d}\psi_{h'_i} \in (x_i - \delta, x_i + \delta)\}.$$

Then, $O_i(x_i, k; \xi, \delta)$ clearly contains h_i; moreover, it is open (Billingsley, 1968, App. III). By choosing $\eta^*, \xi^*, \delta^* > 0$ sufficiently close to 0, we can ensure that both actions are strictly rationalizable for the types in $O_i(x_i, k; \eta^*, \xi^*, \delta^*)$. Let $O_{x_i,k} := O_i(x_i, k; \eta^*, \xi^*, \delta^*)$. Again, multiplicity is robust for any $\varepsilon \in [0, \frac{1}{2})$. □

D.8 Proof of Theorem 4.7

Each type in T^ε is characterized completely by its signal x_i and its reasoning ability (i.e., its depth of reasoning and higher-order beliefs about players' depth of reasoning). We can quantify the latter by assigning a rank to each type. For $n < \infty$, the rank of a type $h_i \in T^{\varepsilon,n}_i$ is just its depth n. The rank of the other types can be assigned using transfinite ordinals. Define the rank of a type h_i in $T^{\varepsilon,\infty}_i$ to be ω (where ω is the first countable ordinal), and for $n > 0$, let the rank of a type in $T^{\varepsilon,\infty+n}$ be $\omega + n$. Types with the same rank have the same depth of reasoning and the same higher-order beliefs about depth of reasoning.

By Lemma 4.6, for every finite α and any $\varepsilon \in [0, \frac{1}{2})$, there is $\underline{x}^\varepsilon_\alpha \in (0, \frac{1}{2})$ and $\bar{x}^\varepsilon_\alpha \in (\frac{1}{2}, 1)$ such that for every type h_i in T^ε with rank α and signal $x_i \in (\underline{x}^\varepsilon_\alpha, \bar{x}^\varepsilon_\alpha)$, both actions are robustly rationalizable for h_i.

Let $\alpha = \omega$, and fix $z \in (\frac{1}{2}, 1)$. Then, there is finite k_z such that any type in T^ε with rank α assigns probability less than z to types with rank greater than k_z. By construction, k_z does not depend on ε.

For ease of notation, write $\underline{x} := \underline{x}^\varepsilon_{k_z}$ and $\bar{x} := \bar{x}^\varepsilon_{k_z}$. Fix $\tilde{x}_i \in (0, \frac{1}{2})$ such that $\tilde{x}_i > 1 - z, \underline{x}$ and hence $\tilde{x}_i < z, \bar{x}$ (as $z, \bar{x} > \frac{1}{2}$). (Such a signal $\tilde{x}_i$ exists, since $1 - z, \underline{x} < \frac{1}{2}$.)

Then, there is $\varepsilon_{\tilde{x}_i} > 0$ such that if $\varepsilon \leq \varepsilon_{\tilde{x}_i}$, a type in T^ε with rank α and signal $y_i \in [\tilde{x}_i, \frac{1}{2}]$ assigns probability 1 to the opponent having a signal in $(\underline{x}, \bar{x})$. If the type conjectures that the opponent invests whenever his rank is $k \leq k_z$ and his signal is in $(\underline{x}^\varepsilon_k, \bar{x}^\varepsilon_k) \supset (\underline{x}, \bar{x})$, then

the expected payoff to I is at least $(1-z)\cdot y_i + z\cdot(y_i-1) > 0$. Under the conjecture that the opponent does not invest whenever his rank is $k \le k_z$ and his signal is in $(\underline{x}^\varepsilon_k, \bar{x}^\varepsilon_k) \supset (\underline{x}, \bar{x})$, the expected payoff to I is less than $z\cdot\frac{1}{2} - (1-z)\cdot\frac{1}{2} = 0$. So, if $\varepsilon < \varepsilon_{x_i}$, then both actions are strictly rationalizable for a type in T^ε with rank α and signal $y_i \in [\tilde{x}_i, \frac{1}{2}]$. Likewise, for $\tilde{x}'_i \in (\frac{1}{2}, 1)$, there is $\varepsilon_{\tilde{x}'_i} > 0$ such that both actions are strictly rationalizable for a type in T^ε with rank α and signal $y_i \in [\frac{1}{2}, \tilde{x}'_i]$. Define $\underline{x}_\alpha := \tilde{x}_i$, $\bar{x}_\alpha := \tilde{x}'_i$, and $\tilde{\varepsilon} := \min\{\varepsilon_{\tilde{x}_i}, \varepsilon_{\tilde{x}'_i}\}$.

For $n > 0$, suppose that for $\gamma = \omega, \omega+1, \ldots, \omega+n-1$, there exist $\varepsilon_\gamma > 0$, $\underline{x}^\varepsilon_\gamma < \frac{1}{2}$, $\bar{x}^\varepsilon_\gamma > \frac{1}{2}$ such that for any rank-γ type in T^ε both actions are strictly rationalizable whenever its signal lies strictly between $\underline{x}^\varepsilon_\gamma$ and $\bar{x}^\varepsilon_\gamma$ and $\varepsilon < \varepsilon_\gamma$. Fix $\tilde{x}_i \in (x^\varepsilon_{\omega+n-1}, \frac{1}{2})$. Then, there is $\varepsilon_{\tilde{x}_i} > 0$ such that if $\varepsilon < \varepsilon_{\tilde{x}_i}$, a type in T^ε assigns probability 1 to the opponent having a type in $(\underline{x}^\varepsilon_{\omega+n-1}, \bar{x}^\varepsilon_{\omega+n-1})$. So, if $\varepsilon < \min\{\varepsilon_\omega, \ldots, \omega_{\omega+n-1}, \varepsilon_{\tilde{x}_i}\}$, both actions are strictly rationalizable for a rank-$(\omega+n)$ type in T^ε with signal $\tilde{x}_i$. By a similar argument, for a fixed $\tilde{x}'_i \in (\frac{1}{2}, x^\varepsilon_{\omega+n-1})$, we can find $\varepsilon_{\tilde{x}'_i} > 0$ such that both actions are strictly rationalizable for any rank-$(\omega+n)$ type in T^ε with signal $\tilde{x}'_i$ whenever $\varepsilon < \min\{\varepsilon_\omega, \ldots, \omega_{\omega+n-1}, \varepsilon_{\tilde{x}'_i}\}$. Define $\underline{x}^\varepsilon_\alpha := \tilde{x}_i$, $\bar{x}^\varepsilon_\alpha := \tilde{x}'_i$, and $\varepsilon_{\omega+n} := \min\{\varepsilon_\omega, \ldots, \omega_{\omega+n-1}, \varepsilon_{\tilde{x}_i}, \varepsilon_{\tilde{x}'_i}\}$.

So, for every infinite rank α, there exist $\varepsilon_\alpha > 0$ and bounds $\underline{x}^\varepsilon_\alpha < \frac{1}{2}$, $\bar{x}^\varepsilon_\alpha > \frac{1}{2}$ such that both actions are strictly rationalizable for a rank-α type in T^ε with a signal strictly between these bounds whenever $\varepsilon < \varepsilon_\alpha$. We next show that this multiplicity is robust. Recall that by Lemma 4.6, for every finite rank γ, there exist $\underline{x}^\varepsilon_\gamma \in (0, \frac{1}{2})$ and $\bar{x}^\varepsilon_\gamma \in (\frac{1}{2}, 1)$ such that every rank-γ type in T^ε with signal $x_i \in (\underline{x}^\varepsilon_\gamma, \bar{x}^\varepsilon_\gamma)$ has a neighborhood $O_{x_i,\gamma}$ such that both actions are rationalizable for the types in $O_{x_i,\gamma}$.

Suppose $\varepsilon < \varepsilon_\omega$, and fix a type h_i in T^ε with rank ω and signal $x_i \in (\underline{x}^\varepsilon_\omega, \bar{x}^\varepsilon_\omega)$. By construction, h_i assigns probability 1 to the opponent having a signal in $(\underline{x}^\varepsilon_{k_z}, \bar{x}^\varepsilon_{k_z})$ and probability less than z to types with rank greater than k_z. Since ψ_{h_i} is a regular probability measure, for every $\eta > 0$, there is a compact subset K_η of types that have rank at most k_z and whose signal is in $(\underline{x}^\varepsilon_{k_z}, \bar{x}^\varepsilon_{k_z})$ such that $\psi_{h_i}(K_\eta) > 1 - z - \eta$. Moreover, K_η has a (finite) open cover $V_{x_i,\omega,\eta} := \bigcup_{m=1}^{\ell} O_{x_i^m,\gamma^m}$, where $\gamma^m \le k_z$ and $x_i^m \in (\underline{x}^\varepsilon_{\gamma^m}, \bar{x}^\varepsilon_{\gamma^m}))$. Then, by a similar argument as before, we can construct a neighborhood $O_{x_i,\omega}$ of h_i such that both actions are strictly rationalizable across all types in the neighborhood. The proof for $\alpha = \omega+1, \omega+2, \ldots$ is now straightforward, and thus omitted. □

D.9 Proof of Proposition 5.1

Let $\hat{H}_i$ be the set of types from finite models.[24] Fix a finite model $M = (\widetilde{\Theta}, T)$ consistent with common belief in an infinite depth, and let $h_i^* \in T_i$ be a type with multiple rationalizable

[24] That is, $h_i \in \hat{H}_i$ if and only if there is a finite model $M'' = (\widetilde{\Theta}'', T'')$ such that $h_i \in T''_i$.

actions. That is, $R_i^\infty(h_i^*) \supset \{a, b\}$ for two distinct actions $a, b \in A_i$. Fix $\eta > 0$ and $\kappa < \infty$. By Corollary 2 of Weinstein and Yildiz (2007), every (η, κ)-ball $O_{\eta,\kappa}(h_i^*)$ of h_i^* has a nonempty intersection with the sets $U_i^a := \{h_i \in \hat{H}_i : R_i^\infty(h_i) = \{a\}\}$ and $U_i^b := \{h_i \in \hat{H}_i : R_i^\infty(h_i) = \{b\}\}$ of types for whom a and b, respectively, are the unique rationalizable action. Let $h_i^a \in U_i^a \cap O_\varepsilon(h_i^*)$ and $h_i^b \in U_i^b \cap O_\varepsilon(h_i^*)$, so $R_i^\infty(h_i^a) = \{a\}$ and $R_i^\infty(h_i^b) = \{b\}$.

Define the model (M^e, τ^e), $e = a, b$, as follows. For $e = a, b$, let $\tilde{M}^e = (\widetilde{\Theta}^e, \tilde{T}^e)$ be a finite model that contains h_i^e. (Such a model exists, since $h_i^e \in \hat{H}_i$.) Then, let $M^e = (\widetilde{\Theta} \cup \widetilde{\Theta}^e, T^e)$, where $T_i^e := T_i \cup \tilde{T}_i^e$ for $i = 1, 2$. Define $\tau^e : T \to T^e$ by $\tau_i^e(h_i^*) := h_i^e$, and $\tau_j^e(h_j) = h_j$ for $h_j \neq h_i^*$. Then, (M^a, τ^a) and (M^b, τ^b) are (η, κ)-perturbations of M.

If an equilibrium σ is (η, κ)-robust, then there is a Bayesian-Nash equilibrium σ^a for M^a and a Bayesian-Nash equilibrium σ^b for M^b such that

$$\sigma_i^a(h_i^a) = \sigma_i^a(\tau_i^a(h_i^*)) = \sigma_i(h_i^*) = \sigma_i^b(\tau_i^b(h_i^*)) = \sigma_i^b(h_i^b).$$

But, as equilibrium refines rationalizability, $\sigma_i^a(h_i^a) = a$ and $\sigma_i^b(h_i^b)$. So, σ is not (η, κ)-robust. Since a similar argument holds for any Bayesian-Nash equilibrium and $\eta > 0$, $\kappa < \infty$, there is no robust equilibrium for M. □

D.10 Proof of Theorem 5.2

As noted in the main text, we take X_i to be a singleton here. It should be clear how to extend the result to the general case.

We construct a set of models, one model $M^{A'}$ with beliefs in the multiplicity set $\Delta^{A'}$, and one model M^a for each action profile a' (possibly empty). We then use these models to define a larger model M. Considering the models M^a is not necessary to prove the result. However, it will be instructive to compare the robustness proof for the types with multiplicity (in $M^{A'}$) with the proof for the types with a dominant action (in M^a).

Step 1. Defining the model $M^{A'}$. Fix $k^{A',<\infty} = 0, 1, \ldots$ and $k^{A',\infty} = 1, 2, \ldots$. Let $T_i^{A',0} := H_i^0$, and for $m = 1, \ldots, k^{A',<\infty}$, let $T_i^{A',m}$ be a finite set of depth-m types with belief in $\Delta_i^{A'}$ that assign probability 1 to $T_{-i}^{A',m-1}$. For $m = 1, \ldots, k^{A',\infty}$, let $T_i^{A',k^{<\infty}+m}$ be a finite set of infinite-depth types with belief in $\Delta_i^{A'}$ that assign probability 1 to $\bigcup_{\ell=1}^{k^{A',<\infty}+m-1} T_{-i}^{A',\ell}$.

Let $T_i^{A'} := \bigcup_{\ell=0}^{k^{A',<\infty}+k^{A,\infty}} T_i^{A',\ell}$. Then, $M^{A'} := (\widetilde{\Theta}^{A'}, T^{A'})$ is a finite model. Every type in $T_i^{A',m}$ assigns probability 1 to types in $\bigcup_{\ell<m} T_{-i}^{A',\ell}$. Say that the *type-rank* of a type $t_i \in T_i^{A',m}$ is m. By construction, each type in $M^{A'}$ has a unique type-rank. The model $M^{A'}$ has level-k beliefs; however, this is immaterial for our results.

Step 2. Defining the model M^a for $a \in A$. Fix an action profile $a \in A$. If there is a player i for whom there is no state at which a_i is strictly dominant, then we define $T^a = \emptyset$, $\widetilde{\Theta}^a = \emptyset$. (Of course, this is ruled out if Assumption R-Dom is satisfied.) Otherwise, for $i = 1, 2$, let θ^{a_i} be a state for which a_i is strictly dominant for i. Fix $k^{a,<\infty} = 0, 1, \ldots$ and $k^{a,\infty} = 1, 2, \ldots$. Let $T_i^{a,0} := H_i^0$, and for $m = 1, \ldots, k^{a,<\infty}$, let $T_i^{a,m} = \{t_i^{a,m}\}$ be the depth-m type that assigns probability 1 to θ^{a_i} and to $T_{-i}^{a,m-1}$. For $m = 1, \ldots, k^{a,\infty}$, let $T_i^{a,k^{<\infty}+m}$ be a finite set of infinite-depth types that assign probability 1 to θ^{a_i} and to $\bigcup_{\ell=1}^{k^{a,<\infty}+m-1} T_{-i}^{a,\ell}$. Let $T_i^a := \bigcup_{\ell=0}^{k^{a,<\infty}+k^{a,\infty}} T_i^{a,\ell}$. Then, $M^a := (\widetilde{\Theta}^a, T^a)$ is a finite model. Again, each type in M^a has a unique type-rank.

Step 3. Defining the model M. Let $\widetilde{\Theta} := \widetilde{\Theta}^{A'} \cup \bigcup_a \widetilde{\Theta}^a$ and $T_i := T_i^{A'} \cup \bigcup_a T_i^a$. Then, $M := (\widetilde{\Theta}, T)$ is a finite (nonempty) model.

Step 4. Robustness for $M^{A'}$. We next show that every strict Bayesian-Nash equilibrium of $M^{A'}$ is robust. That is, let σ be a strict Bayesian-Nash equilibrium of $M^{A'}$. Then, we claim that there is $\eta > 0$ and $\kappa < \infty$ such that for every (η, κ)-perturbation (M', τ) of $M^{A'}$, where $M' = (\widetilde{\Theta}', T')$, there is a Bayesian-Nash equilibrium σ' for M' that coincides with σ on $\tau(T^{A'})$.

To show this, let σ be a strict Bayesian-Nash equilibrium for $M^{A'}$. Since σ is a strict Bayesian-Nash equilibrium for $M^{A'}$, there is $z > 0$ such that for every $i = 1, 2$, $t_i \in T_i^{A'}$ and $a_i' \neq \sigma_i(t_i)$,

$$\int_{\Theta \times T_{-i}^{A'}} u_i(\sigma_i(t_i), \sigma_{-i}(t_{-i}), \theta) \mathrm{d}\psi_{t_i} - \int_{\Theta \times T_{-i}^{A'}} u_i(\sigma_i(t_i), \sigma_{-i}(t_{-i}), \theta) \mathrm{d}\psi_{t_i} \geq z.$$

Also note that since u_i is a continuous function defined on a compact space, there is $c > 0$ such that $u_i(a_i, a_{-i}, \theta) \in [-\frac{c}{2}, \frac{c}{2}]$ for all i, a_i, a_{-i}, and θ.

Let $\kappa = k^{A',<\infty} + k^{A',\infty} + 1$. Fix $\eta \in (0, 1)$ such that the (η, κ)-balls of the types in $M^{A'}$ are disjoint (i.e., if t_i, t_i' are distinct types in $T_i^{A'}$, then $O_{\eta,\kappa}(t_i) \cap O_{\eta,\kappa}(t_i') = \emptyset$). (Such an η exists since $M^{A'}$ is finite.) Since $\eta < 1$, if t_i has a finite depth, then any type in $O_{\eta,\kappa}(t_i)$ has the same depth as t_i (and thus the same type-rank).

Let (M', τ), with $M' = (\widetilde{\Theta}', T')$, be an (η, κ)-perturbation of $M^{A'}$. We define auxiliary profiles σ'' and σ'''. Let T_i'' be the set of types $t_i' \in T_i'$ such that there is $t_i \in T_i^{A'}$ such that $t_i' = \tau_i(t_i)$, $t_i' = t_i$, or $t_i' \in O_{\eta,\kappa}(t_i)$. (Note that these possibilities are not mutually exclusive.) Define the strategy profile σ'' for the types in T'' as follows. For every $t_i' \in T_i'$ such that $t_i' = \tau_i(t_i)$ for some $t_i \in T_i^{A'}$, let $\sigma_i''(t_i') = \sigma_i(t_i)$. Otherwise, if $t_i' \in T_i' \cap T_i^{A'}$, let $\sigma_i''(t_i') = \sigma_i(t_i')$. Otherwise, if $t_i' \in O_{\eta,\kappa}(t_i)$ for some $t_i \in T_i^{A'}$, then let $\sigma_i''(t_i') = \sigma_i(t_i)$. (By our choice of η, there is a unique such type t_i.)

Let $T_i''' := T_i' \setminus T_i''$, and for $i = 1, 2$ and $t_i' \in T_i'''$, let $\sigma_i'''(t_i')$ be a best response to the belief that the opponent plays according to σ_{-i}'' if his type is in T_i'' and according to σ_{-i}''' otherwise. (Such a σ''' exists by standard equilibrium existence arguments.)

Then, define the strategy profile σ' as follows: let $\sigma_i'(t_i') = \sigma_i''(t_i')$ if $t_i' \in T_i''$ and $\sigma_i'(t_i') = \sigma_i'''(t_i')$ otherwise. So, $\sigma_i'(t_i')$ is derived from σ_i if t_i' is close to a type in T_i, and is a best response to σ'_{-i} otherwise. Moreover, since $M^{A'}$ is a model and σ is a (strict) Bayesian-Nash equilibrium for $M^{A'}$, the types in $T \cap T'$ also play a best response under σ'. (This follows from a standard "pull-back" property; see Friedenberg and Meier (2016).) This means that to check whether σ' is a Bayesian-Nash equilibrium for M', we need only to check the types in $T'' \setminus T$. These are precisely the types in T' that are in the (η, κ)-neighborhoods of the types in $T^{A'}$ but not in $T^{A'}$ itself.

Consider a type $t_i' \in T_i'' \setminus T_i^{A'}$ of type-rank 1. So, there is $t_i \in T_i^{A'}$ with $\sigma_i'(t_i') = \sigma_i(t_i)$. Moreover, t_i' assigns probability 1 to $\tau(T_{-i}^{A',0}) = T_{-i}^{A',0}$. By construction, $\sigma'_{-i}(t_{-i}) = \sigma_{-i}(t_{-i})$ for $t_{-i} \in T_{-i}^{A',0}$. Also, t_i''s belief about Θ is η-close to t_i' belief about Θ. So, for every $a_i \in A_i$,

$$\Big| \int_{\widetilde{\Theta}' \times T'_{-i}} u_i(a_i, \sigma'_{-i}(t'_{-i}), \theta') \mathrm{d}\psi t_i' - \sum_{\widetilde{\Theta}^{A'} \times T_{-i}^{A'}} u_i(a_i, \sigma_{-i}(t_{-i}), \theta) \mathrm{d}\psi_{t_i} \Big| < \eta c.$$

So, for every $a_i \in A_i$,

$$\int_{\widetilde{\Theta}' \times T'_{-i}} u_i(\sigma_i'(t_i'), \sigma'_{-i}(t'_{-i}), \theta') \mathrm{d}\psi_{t_i'} - \int_{\widetilde{\Theta}' \times T'_{-i}} u_i(a_i, \sigma'_{-i}(t'_{-i}), \theta') \mathrm{d}\psi_{t_i'} \geq z - 2\eta c.$$

Next, consider a type $t_i' \in T_i'' \setminus T_i^{A'}$ of type-rank $m = 2, \ldots, k^{A',<\infty}$. As before, there is $t_i \in T_i^{A'}$ with $\sigma_i'(t_i') = \sigma_i(t_i)$. Fix $a_i \in A_i$. Since

$$\int_{\widetilde{\Theta}' \times T'_{-i}} u_i(a_i, \sigma'_{-i}(t'_{-i}), \theta') \mathrm{d}\psi_{t_i'} = \int_{\widetilde{\Theta}' \times T''_{-i}} u_i(a_i, \sigma'_{-i}(t''_{-i}), \theta') \mathrm{d}\psi_{t_i'} + \int_{\widetilde{\Theta}' \times T'''_{-i}} u_i(a_i, \sigma'_{-i}(t'''_{-i}), \theta') \mathrm{d}\psi_{t_i'},$$

and because for every type $t''_{-i} \in T''_{-i}$, there is $t_{-i} \in T_{-i}$ with $t''_{-i} \in O_{\eta,\kappa}(t_{-i})$ and $\sigma'_{-i}(t''_{-i}) = \sigma_{-i}(t_{-i})$, we have

$$\Big| \int_{\widetilde{\Theta}' \times T'_{-i}} u_i(a_i, \sigma'_{-i}(t'_{-i}), \theta') \mathrm{d}\psi_{t_i'} - \int_{\widetilde{\Theta}^{A'} \times T_{-i}^{A'}} u_i(a_i, \sigma_{-i}(t_{-i}), \theta) \mathrm{d}\psi_{t_i} \Big| < (1 - \eta) \cdot \eta c + \eta c.$$

Hence, for every $a_i \in A_i$,

$$\int_{\widetilde{\Theta}' \times T'_{-i}} u_i(\sigma_i'(t_i'), \sigma'_{-i}(t'_{-i}), \theta') \mathrm{d}\psi_{t_i'} - \int_{\widetilde{\Theta}' \times T'_{-i}} u_i(a_i, \sigma'_{-i}(t'_{-i}), \theta') \mathrm{d}\psi_{t_i'} \geq z - 2\eta c \cdot (2 - \eta).$$

We next consider the (η,κ)-perturbations of the infinite-depth types in $M^{A'}$, that is, the types $t_i' \in T_i'' \setminus T_i^{A'}$ such that there is $t_i \in T_i$ with type-rank $m = k^{A',<\infty}+1,\ldots,k^{A',\infty}$. Suppose t_i' is an (η,κ)-perturbation of a type $t_i \in T_i$ with type-rank $k^{A',<\infty}+1$. The critical observation is that while the depth of reasoning of t_i' can be arbitrarily high (in fact, its depth could be infinite), its belief $\psi_{t_i'}$ is largely determined by its $(k^{A',<\infty}+1)$th-order belief: since t_i assigns probability 1 at order $k^{A',<\infty}+1$ to the opponent having a type in $\bigcup_{\ell=1}^{k^{A',<\infty}} T_{-i}^{A',\ell}$ and t_i' is in the (η,κ)-neighborhood of t_i (for $\kappa \geq k^{A',<\infty}+1$), t_i' assigns probability $1-\eta$ at order $k^{A',<\infty}+1$ to the (η,κ)-neighborhood of $\bigcup_{\ell=1}^{k^{A',<\infty}} T_{-i}^{A',\ell}$. So, mass $1-\eta$ of $\psi_{t_i'}$ is determined at order $k^{A',<\infty}$. A similar argument applies for infinite-depth types with type-rank $m > k^{A',<\infty}+1$ (given that $\kappa = k^{A',<\infty}+k^{A',\infty}+1$).

Hence, the argument for the infinite-depth case is the same as for the finite-depth case: for every $m = k^{A',<\infty}+1,\ldots,k^{A',\infty}$, every type in $t_i' \in T_i'' \setminus T_i^{A'}$ that is an (η,κ)-perturbation of a type in $T_i^{A'}$ with type-rank m, and for every $a_i \in A_i$.

$$\int_{\widetilde{\Theta}'\times T_{-i}'} u_i(\sigma_i'(t_i'),\sigma_{-i}'(t_{-i}'),\theta')\mathrm{d}\psi_{t_i'} - \int_{\widetilde{\Theta}'\times T_{-i}'} u_i(a_i,\sigma_{-i}'(t_{-i}'),\theta')\mathrm{d}\psi_{t_i'} \geq z - 2\eta c\cdot(2-\eta).$$

So, if $\eta > 0$ is sufficiently small so that $z > 2\eta c\cdot(2-\eta)$ and the (η,κ)-balls around the types in M are disjoint, then σ' is a Bayesian-Nash equilibrium. This conclusion does not depend on the particular (η,κ)-perturbation that we considered: for any $\eta > 0$ such that $z > 2\eta c\cdot(2-\eta)$ and the (η,κ)-balls around the types in M are disjoint, any (η,κ)-perturbation has a Bayesian-Nash equilibrium that coincides with σ on the image of $M^{A'}$. So, if $z > 2\eta c\cdot(2-\eta)$, then σ is (η,κ)-robust.

Step 4. Robustness for M^a, $a \in A$. For any nonempty model M^a, every strict Bayesian-Nash equilibrium of M^a is robust. While this can be shown using an argument similar to the one in Step 3, there is a much easier and much more direct proof: every type in M^a has a strictly dominant action; so, any type whose beliefs about Θ are η-close to one of the types in M^a has a unique rationalizable action for $\eta > 0$ sufficiently small (under the usual weak topology on $\Delta(\Theta)$). In this case, the beliefs of types about the opponent's action, or their beliefs about their opponent's belief about their opponent's actions, and so on, are all immaterial.

Step 5. Robustness for M. The proof that any strict Bayesian-Nash equilibrium for M is robust is essentially a combination of Steps 3 and 4, and is thus omitted. (It does not immediately follow from the proofs in Steps 3 and 4 in isolation, however. This is because even though $M^{A'}$ and M^a are disjoint, some of their perturbations may not be (even if η is

small and κ is large): indeed, as perturbed models relax the common-knowledge restrictions embodied in the original model, they tend to be large.) $\square$

The model M is consistent with $(k^{A',\infty}-1)$th-order belief in an infinite depth. So, by choosing $k^{A',\infty}$ appropriately, we obtain a model that is consistent with mth-order belief in an infinite depth for arbitrary depth. $\square$

It is easy to extend the construction: the result would also hold if we had added types to M that assign positive probability to $M^{A'}$ and to M^a, $a \in A$ (assuming the latter is nonempty for some a), as long as incentives remain strict, and the equilibrium actions of types with multiplicity are pinned down by the equilibrium actions of types with a lower level of sophistication (as given by their type-rank).

D.11 Proof of Proposition 5.3

Take $\widetilde{\Theta} := \{\theta^*, \theta^{\tilde{a}}, \theta^{\tilde{b}}\}$, where θ^* is a state at which the complete-information game has two strict Nash equilibria, and $\theta^{\tilde{a}}$ and $\theta^{\tilde{b}}$ are states at which $\tilde{a}$ and $\tilde{b}$ are strictly dominant for both players. (Such states exist by the definition of global games and the assumption of symmetry.)

Define the model $M = (\widetilde{\Theta}, T)$ as in the proof of Theorem 5.2, except that the types in $T_i^{A',m}$ assign probability 1 to θ^* and that we take $T_a = \emptyset$ if $a \neq (\tilde{a}, \tilde{a}), (\tilde{b}, \tilde{b})$. Then, M is a finite model with complete information.

Define the strategy profile $\sigma^{\tilde{a}}$ as follows. For any type h_i in $M^{(\tilde{a},\tilde{a})}$ (where $\tilde{a}$ is strictly dominant) or in $M^{A'}$ (where the complete-information game has two strict Nash equilibria), define $\sigma_i^{\tilde{a}}(h_i) = \tilde{a}$. For any type h_i in $M^{(\tilde{b},\tilde{b})}$ (where $\tilde{b}$ is strictly dominant), define $\sigma_i^{\tilde{a}}(h_i) = \tilde{b}$.

Define $\sigma^{\tilde{b}}$ analogously: for any type h_i in $M^{(\tilde{b},\tilde{b})}$ (where $\tilde{b}$ is strictly dominant) or in $M^{A'}$ (where the complete-information game has two strict Nash equilibria), define $\sigma_i^{\tilde{b}}(h_i) = \tilde{b}$. For any type h_i in $M^{(\tilde{a},\tilde{a})}$ (where $\tilde{a}$ is strictly dominant), define $\sigma_i^{\tilde{b}}(h_i) = \tilde{a}$.

It is easy to check that $\sigma^{\tilde{a}}$ and $\sigma^{\tilde{b}}$ are strict Bayesian-Nash equilibria. As such, they are robust (Theorem 5.2). In particular, the predictions remain valid if we introduce a small amount of information about payoffs. $\square$

D.12 Proof of Lemma A.1

Clearly, $R_i^{T,0}(t_i) = R_i^0(h_i^T(t_i))$. For $m > 0$, suppose that for all $n \leq m-1$, $R_i^{T,n} = R_i^n \circ h_i^T$. As in the proof of Lemma D.2, if $a_i \in R_i^m(h_i^T(t_i))$, then there is a measurable conjecture $s_{-i} : \Theta \times H_{-i} \to \Delta(A_{-i})$ such that a_i is a best response for $h_i^T(t_i)$ under the conjecture s_{-i}. Then, $s_{-i} \circ h_{-i}^T$ is a measurable conjecture such that a_i is a best response for t_i under the conjecture, so $a_i \in R_i^{T,m}(t_i)$. Conversely, suppose $a_i \in R_i^{T,m}(t_i)$. Then there is a belief $\mu_{t_i} \in \Delta(\Theta \times \mathrm{Gr}(R_{-i}^{T,m-1}))$ so that a_i is a best response to μ_{t_i}. The belief μ_{t_i}

defines a belief $\mu_{h_i} \in \Delta(\Theta \times \mathrm{Gr}(R_{-i}^{m-1}))$ in the obvious way. Then, by the Kuratowski-Ryll-Nardzweski selection theorem, there is a measurable conjecture $s_{-i} : \Theta \times H_{-i} \to \Delta(A_{-i})$ such that a_i is a best response against s_{-i} for $h_i^T(t_i)$, so $a_i \in R_i^m(h_i^T(t_i))$. It is now immediate that $R_i^{T,\infty}(t_i) = R_i^\infty(h_i^T(t_i))$. □

D.13 Proof of Lemma B.1

The proof follows from a number of lemmas:

Lemma D.4. *For $i = 1, 2$ and $k \in \mathbb{N}$, $\widetilde{\Omega}_i^k$, Ω_i^k, $\widetilde{H}_i^k$ and H_i^k are compact metric.*

Proof. The proof is by induction. Clearly, $\widetilde{H}_i^0$ and H_i^0 are compact metric, so that $\widetilde{\Omega}_i^0$, Ω_i^0 and $\widetilde{H}_i^1$ and H_i^1 are also compact metric. Suppose $\widetilde{\Omega}_i^\ell$, Ω_i^ℓ, $\widetilde{H}_i^{\ell+1}$ and $H_i^{\ell+1}$ are compact metric for each $i = 1, 2$ and $\ell \leq k-1$. Then, $\widetilde{\Omega}_i^k$ and Ω_i^k are compact metric. It remains to show that $\widetilde{H}_i^{k+1}$ and H_i^{k+1} are compact metric. As $\Delta(\widetilde{\Omega}_i^k)$ and $\Delta(\Omega_i^k)$ are compact metric, we need to show that $\widetilde{H}_i^{k+1}$ and H_i^{k+1} are a closed subset of $\widetilde{H}_i^k \times \Delta(\widetilde{\Omega}_i^k)$ and $\widetilde{H}_i^k \times \Delta(\Omega_i^k)$, respectively. We prove the claim for $\widetilde{H}_i^{k+1}$; the proof for H_i^{k+1} is similar. Let $h_i = (x_i, \mu_i^0, \ldots, \mu_i^{k+1}) \in \widetilde{H}_i^k \times \Delta(\widetilde{\Omega}_i^k)$ and suppose there is a sequence $(h_i^n)_{n\in\mathbb{N}}$ in $\widetilde{H}_i^{k+1}$, where $h_i^n = (x_i^n, \mu_i^{0,n}, \mu_i^{2,n}, \ldots, \mu_i^{k+1,n})$, such that $h_i^n \to h_i$. It is sufficient to show that $h_i \in \widetilde{H}_i^k$. If we show that

$$\mathrm{marg}_{\widetilde{\Omega}_i^{k-1}} \mu_i^{k+1,n} \to \mathrm{marg}_{\widetilde{\Omega}_i^{k-1}} \mu_i^{k+1}, \tag{D.1}$$

and

$$\mu_i^{k,n} \to \mu_i^k, \tag{D.2}$$

the proof is complete: Because $h_i^n \in \widetilde{H}_i^{k+1}$ for all n, it follows that

$$\mathrm{marg}_{\widetilde{\Omega}_i^{k-1}} \mu_i^{k+1} = \mu_i^k,$$

so that $h_i \in \widetilde{H}_i^{k+1}$. But using that $\widetilde{H}_i^k \times \Delta(\widetilde{\Omega}_i^k)$ is endowed with the product topology, (D.1) and (D.2) follow immediately from the assumption that $h_i^n \to h_i$. □

Lemma D.5. (Heifetz, 1993, Thm. 6) *For any $(x_i, \mu_i^0, \ldots, \mu_i^k) \in \widetilde{H}_i^k$, there exists $\mu_i^{k+1} \in \Delta(\widetilde{\Omega}_i^k)$ such that $(x_i, \mu_i^0, \ldots, \mu_i^k, \mu_i^{k+1}) \in \widetilde{H}_i^{k+1}$.*

The proof is similar to the proof of Theorem 6 of Heifetz (1993) and thus omitted. We are now ready to prove Lemma B.1. By Lemma D.5, $\widetilde{H}_i^k$ is nonempty. Also, the projection function from $\widetilde{H}_i^k$ into $\widetilde{H}_i^{k-1}$ is surjective. By standard arguments, the inverse limit space H_i^∞ is nonempty. Since H_i^∞ is a closed subset of the compact metric space $\widetilde{H}_i^0 \times \prod_{k=0}^\infty \Delta(\Omega_i^k)$, it is compact metric. Finally, H_i is Polish since it is the disjoint union of a countable family of compact metric (and thus Polish) spaces. □

D.14 Proof of Lemma B.2

We first prove the first claim. By Lemma B.1, the space $\Theta \times H_{-i}^{\infty}$ is a nonempty Polish space for every player i. By a version of the Kolmogorov consistency theorem, for each belief hierarchy $h_i^{\infty} = (x_i, \mu_i^0, \mu_i^1, \ldots) \in H_i^{\infty}$ of infinite depth, there exists a unique Borel probability measure μ_i^{∞} on $\Theta \times H_{-i}$ such that

$$\text{marg}_{\widetilde{\Omega}_i^k} \mu_i^{\infty} = \mu_i^{k+1}$$

for all k, i.e., the mapping is canonical. The last claim follows immediately by associating the belief μ_i^k to the finite hierarchy $h_i^k = (x_i, \mu_i^0, \ldots, \mu_i^{k-1}, \mu_i^k) \in \widetilde{H}_i^k$. □

D.15 Proof of Proposition B.3

First consider the infinite-depth hierarchies. Lemma B.2 shows that each infinite belief hierarchy $h_i^{\infty} = (x_i, \mu_i^0, \mu_i^1, \ldots) \in H_i^{\infty}$ corresponds to a unique Borel probability measure on $\Theta \times H_{-i}$, and the mapping is canonical. Moreover, the signal x_i associated with h_i^{∞} is obtained by projecting h_i^{∞} onto X_i. Denote the function that maps H_i^{∞} into $X_i \times \Delta(\Theta \times H_{-i})$ in this way by $\tilde{\psi}_i^{\infty}$. Conversely, let $r_i^{\infty} : X_i \times \Delta(\Theta \times H_{-i}) \to H_i^{\infty}$ be the mapping that assigns to each $(x_i, \mu_i) \in X_i \times \Delta(\Theta \times H_{-i})$ the hierarchy $(x_i, \text{marg}_{\Theta}\mu_i, \text{marg}_{\widetilde{\Omega}_i^0}\mu_i, \text{marg}_{\widetilde{\Omega}_i^1}\mu_i, \ldots) \in X_i \times \Delta(\Theta) \times \prod_{k \geq 0} \Delta(\widetilde{\Omega}_i^k)$. The function r_i^{∞} is the inverse of $\tilde{\psi}_i^{\infty}$; it remains to show that $\tilde{\psi}_i^{\infty}$ and r_i^{∞} are continuous. The function $\tilde{\psi}_i^{\infty}$ is continuous if and only if $h_i^n \to h_i$ in H_i^{∞} implies $\psi_i^{\infty}(h_i^n) \to \psi_i^{\infty}(h_i)$ in $X_i \times \Delta(\Theta \times H_{-i})$. This follows from the continuity of the projection function and the fact that the cylinders form a convergence-determining class in $\Theta \times H_{-i}$, with the value of $\tilde{\psi}_i^{\infty}(h_i)$ for $h_i = (x_i, \mu_i^0, \mu_i^1, \ldots)$ on the cylinders being given by the μ_i^k's. Finally, it follows from the continuity of the identity function and the marginal operator that r_i^{∞} is continuous.

For the case of finite-depth hierarchies, simply set $\psi_i^k(h_i^k) := (x_i, \mu_i^k)$ for each $h_i^k = (x_i, \mu_i^0, \ldots, \mu_i^{k-1}, \mu_i^k) \in \widetilde{H}_i^k$. Continuity of the mapping ψ_i^k is immediate. □

References

Agranov, M., E. Potamites, A. Schotter, and C. Tergiman (2012). Beliefs and endogenous cognitive levels: An experimental study. *Games and Economic Behavior 75*, 449–463.

Alaoui, L. and A. Penta (2016). Endogenous depth of reasoning. *Review of Economic Studies*. Forthcoming.

Aliprantis, C. D. and K. C. Border (2005). *Infinite Dimensional Analysis: A Hitchhiker's Guide* (3rd ed.). Berlin: Springer.

Angeletos, G.-M., C. Hellwig, and A. Pavan (2006). Signaling in a global game: Coordination and policy traps. *Journal of Political Economy 114*, 452–484.

Angeletos, G.-M., C. Hellwig, and A. Pavan (2007). Dynamic global games of regime change: Learning, multiplicity, and the timing of attacks. *Econometrica 75*, 711–756.

Angeletos, G.-M. and I. Werning (2006). Crises and prices: Information aggregation, multiplicity, and volatility. *American Economic Review 96*, 1720–1736.

Bergemann, D. and S. Morris (2005). Robust mechanism design. *Econometrica 73*, 1771–1813.

Billingsley, P. (1968). *Convergence of Probability Measures*. New York: Wiley.

Brandenburger, A. (2003). On the existence of a 'complete' possibility structure. In M. Basili, N. Dimitri, and I. Gilboa (Eds.), *Cognitive processes and economic behavior*, pp. 30–34. London: Routledge.

Carlsson, H. and E. van Damme (1993). Global games and equilibrium selection. *Econometrica 61*, 989–1018.

Chen, Y.-C. (2012). A structure theorem for rationalizability in the normal form of dynamic games. *Games and Economic Behavior 75*, 587–597.

Chen, Y.-C., A. Di Tillio, E. Faingold, and S. Xiong (2010). Uniform topologies on types. *Theoretical Economics 5*, 445–478.

Chen, Y.-C., A. Di Tillio, E. Faingold, and S. Xiong (2015). The strategic impact of higher-order beliefs. Working paper.

Chen, Y.-C., S. Takahashi, and S. Xiong (2014a). The robust selection of rationalizability. *Journal of Economic Theory 151*, 448–475.

Chen, Y.-C., S. Takahashi, and S. Xiong (2014b). The Weinstein-Yildiz selection and robust predictions with arbitrary payoff uncertainty. Working paper.

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.

De Clippel, G., R. Saran, and R. Serrano (2014). Mechanism design with bounded depth of reasoning and small modeling mistakes. Working paper, Brown University.

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

Dekel, E., D. Fudenberg, and S. Morris (2007). Interim correlated rationalizability. *Theoretical Economics 2*, 15–40.

Dekel, E. and M. Siniscalchi (2014). Epistemic game theory. In P. Young and S. Zamir (Eds.), *Handbook of Game Theory*, Volume 4. Elsevier. Forthcoming.

Friedenberg, A. and M. Meier (2016). The context of the game. *Economic Theory*. Forthcoming.

Harsanyi, J. (1967). Games on incomplete information played by Bayesian players. Part I. *Management Science 14*, 159–182.

Heifetz, A. (1993). The Bayesian formulation of incomplete information–The non-compact case. *International Journal of Game Theory 21*, 329–338.

Heifetz, A. and W. Kets (2011). All types naive and canny. Working paper.

Heifetz, A. and D. Samet (1998). Topology-free typology of beliefs. *Journal of Economic Theory 82*, 324–341.

Kets, W. (2010). Bounded reasoning and higher-order uncertainty. Working paper, Northwestern University.

Kets, W. (2013). Finite depth of reasoning and equilibrium play in games with incomplete information. Working paper, Northwestern University.

Kreps, D., P. Milgrom, J. Roberts, and R. Wilson (1982). Rational cooperation in the finitely-repeated prisoners' dilemma. *Journal of Economic Theory 27*, 245–252.

Lipman, B. L. (1994). A note on the implications of common knowledge of rationality. *Games and Economic Behavior 6*, 114–129.

Lubin, A. (1974). Extensions of measures and the von Neumann selection theorem. *Proceedings of the American Mathematical Society 43*, 118–122.

Mertens, J. F. and S. Zamir (1985). Formulation of Bayesian analysis for games with incomplete information. *International Journal of Game Theory 14*, 1–29.

Milgrom, P. and J. Roberts (1982). Predation, reputation and entry deterrence. *Journal of Economic Theory 27*, 280–312.

Morris, S. and H. S. Shin (2003). Global games: Theory and applications. In M. Dewatripont, L. P. Hansen, and S. J. Turnovsky (Eds.), *Advances in economics and econometrics: Eighth World Congress*, Volume 1. Cambridge University Press.

Penta, A. (2012). Higher order uncertainty and information: Static and dynamic games. *Econometrica 80*, 631–666.

Penta, A. (2013). On the structure of rationalizability for arbitrary spaces of uncertainty. *Theoretical Economics 8*, 405–430.

Purves, R. (1966). Bimeasurable functions. *Fundamenta Mathematicae 58*, 149–157.

Rubinstein, A. (1989). The electronic mail game: Strategic behavior under "almost common knowledge". *American Economic Review 79*, 385–391.

Strzalecki, T. (2014). Depth of reasoning and higher order beliefs. *Journal of Economic Behavior & Organization*. Forthcoming.

Weinstein, J. and M. Yildiz (2007). A structure theorem for rationalizability with application to robust predictions of refinements. *Econometrica 75*, 365–400.

Weinstein, J. and M. Yildiz (2013). Robust predictions in infinite-horizon games–an unrefinable folk theorem. *Review of Economic Studies 80*, 365–394.

Yang, M. (2013). Coordination with flexible information acquisition. Working paper, Duke University.

Yildiz, M. (2005). Generic uniqueness of rationalizable actions. Working paper, MIT.