Forthcoming in Econometrica

A STRUCTURE THEOREM FOR RATIONALIZABILITY WITH APPLICATION TO ROBUST PREDICTIONS OF REFINEMENTS

JONATHAN WEINSTEIN AND MUHAMET YILDIZ

Abstract. Rationalizability is a central solution concept of game theory. Economic models often have many rationalizable outcomes, motivating economists to use refinements of rationalizability, including equilibrium refinements. In this paper we try to achieve a general understanding of when this multiplicity occurs and how we should deal with it. Assuming that the set of possible payoff functions and belief structures is sufficiently rich, we establish a revealing structure of the correspondence of beliefs to sets of rationalizable outcomes. We show that, for any rationalizable action a of any type, we can perturb the beliefs of the type in such a way that a is uniquely rationalizable for the new type. This unique outcome will be robust to further small changes. When multiplicity occurs, then, we are in a "knife-edge" case where the unique rationalizable outcome changes, sandwiched between open sets of types where each of the rationalizable actions is uniquely rationalizable. As an immediate application of this result, we characterize, for any refinement of rationalizability, the predictions that are robust to small misspecifications of interim beliefs. These are only those predictions that are true for all rationalizable strategies, i.e., the predictions that could have been made without the refinement.

Keywords: rationalizability, incomplete information, robustness, refinement, higher-order beliefs, dominance solvability.

The results in this paper have been presented in working papers Weinstein and Yildiz (2004) and Yildiz (2005). We thank Daron Acemoglu, Pierpaolo Battigalli, Adam Brandenburger, Glenn Ellison, Jeffrey Ely, Drew Fudenberg, Aviad Heifetz, Bart Lipman, Casey Rothschild, Dov Samet, and the seminar participants at UAB, Berkeley, Bogazici, Caltech, UCL, Econometric Society Winter Meeting 2004, Econometric Society World Congress 2005, ESSET Meeting, Georgetown, Harvard, Koc, LSE, MIT, Northwestern, NYU, Penn State, UPenn, Sabanci, Stanford, Washington University, Wisconsin-Madison, University of Tokyo, and Yale for helpful comments. We thank Eddie Dekel and three anonymous referees for detailed comments. We are especially grateful to Stephen Morris for extensive discussions that led to numerous improvements and some new results in the paper. The first version was written while Yildiz visited Cowles Foundation of Yale University, and he thanks them for their generous support and hospitality.

1. Introduction

In economic models, there are often many Nash equilibria and a very large number of rationalizable outcomes. In order to be able to make sharp predictions, game theorists therefore developed stronger solution concepts, which led to a multitude of refinements, such as perfect and robust equilibrium. In applications, researchers typically use these refinements, often applying them to Bayesian games in which specific type spaces are chosen to model the players' incomplete information. In this paper, we examine the premises of this program when there is no common-knowledge restriction on payoff functions. Using existing ideas developed in specific contexts, we take a general approach in order to understand when and why there are multiple rationalizable outcomes and how we should address such multiplicity when it occurs. The solution of this problem also allows us to determine which predictions of refinements retain their validity when we actually have only partial knowledge of the players' incomplete information. We show that these are precisely those predictions that are true for all rationalizable strategies, thus voiding the whole point of the refinements (to the extent that our robustness notion is compelling). (Such absence of new robust predictions has been established for many traditional equilibrium refinements by Fudenberg, Kreps, and Levine (1988) and Dekel and Fudenberg (1990), as they have shown that we can make any equilibrium strict by perturbing the payoffs.)

We start with the observation that modeling a given situation inherently involves abstracting away from details and making strong simplifying assumptions that are meant to capture the essence of its true underlying features. In particular, game theoretical models often assume that a particular information structure is common knowledge. These assumptions are meant to be satisfied only approximately in the actual situation. They may nevertheless have a significant impact on the conclusions (see Kreps, Milgrom, Roberts, and Wilson (1982) and, more closely to our paper, Rubinstein (1989)). Carlsson and van Damme (1993) illustrated that multiplicity may sometimes be a direct result of the implicit simplifying assumptions of our models. To be concrete, consider their well-known example.

Example 1 (Carlsson and van Damme (1993)). Consider the payoff matrix

	Attack	No Attack
Attack	θ, θ	$\theta - 1, 0$
No Attack	$0, \theta - 1$	$0, 0$

where θ is a real number. Assume that θ is unknown but each player $i \in \{1, 2\}$ observes a noisy signal $x_i = \theta + \varepsilon\eta_i$, where (η_1, η_2) is distributed independently from θ, and the support of θ contains an interval $[a, b]$ where $a < 0 < 1 < b$. When $\varepsilon = 0$, θ is common knowledge. If it is also the case that $\theta \in (0, 1)$, there exist two Nash equilibria in pure strategies and one Nash equilibrium in mixed strategies. Under mild conditions, Carlsson and van Damme show that when ε is small but positive, multiplicity disappears: when $x_i \neq 1/2$, there is a unique rationalizable action. The rationalizable action is No Attack when $x_i < 1/2$, and it is Attack when $x_i > 1/2$.

The model in which $\theta = \bar{\theta} \neq 1/2$ is common knowledge (i.e. $\varepsilon = 0$) idealizes the situation in which incomplete information is small (i.e. ε is small but positive) and each x_i is close to $\bar{\theta}$, in the following sense. In the interim stage, observing x_i, the player forms beliefs about θ, which are called first-order beliefs, beliefs about the other player's belief about θ, which are called second-order beliefs, and so on. For any k, as ε goes to zero and x_i goes to $\bar{\theta}$, the player's belief at order k converges to the kth-order belief under $\varepsilon = 0$ and $x_i = \bar{\theta}$, i.e., θ is equal to $\bar{\theta}$, the other player knows this, and so on. We, modelers, use the idealized model of $\varepsilon = 0$ as an approximation of this small incomplete information, in order to simplify our model. The simplification weakens our ability to make predictions, as there are now new outcomes that would have been ruled out in a "more complete" model in which our assumptions are relaxed.

Our main result in this paper has two parts. The first part shows that this intuition of Carlsson and van Damme is quite general, using the notion of convergence at each order described in the previous paragraph. Take any situation with multiple rationalizable outcomes. Under their assumption that each action can be dominant at some parameter value, we show that, by introducing a small amount of incomplete information, we can *always* relax the implicit assumptions of the model and obtain an open set of situations in which there is a unique rationalizable outcome, which is the same across all these situations. Therefore, without a very precise knowledge of the actual situation, we cannot rule out the possibility that we could have predicted accurately what the rationalizable outcome is by learning more about the situation. In contrast, when there is a unique rationalizable outcome, slight relaxations of the assumptions will not have any effect.

In Example 1, all the perturbations lead to the same rationalizable strategy as $\varepsilon \to 0$, the strategy in which player takes the risk-dominant action. Carlsson and van Damme have used this observation to argue for the selection of risk-dominant equilibrium in the complete-information game. This particular selection is, however, a property of the specific perturbations considered. Indeed, if we allow the ex ante variance of θ to depend on ε, we can select any outcome we want. This is illustrated in the following example, which is based on Morris and Shin (2000). (Complete details of this example can be found in the appendix.)

Example 2 (Morris and Shin (2000)). In Example 1, assume that θ is normally distributed with mean y and variance $\varepsilon/\sqrt{2\pi}$ and η_i is normally distributed with zero mean and unit variance, where θ, η_1, and η_2 are all independent. Such a prior belief on θ arises if players observe a normally distributed "public" signal y in addition to their "private" signals x_1 and x_2. As $\varepsilon \to 0$, the interim beliefs of any type x_i converge to the complete information game with $\theta = x_i$. Morris and Shin (2000) have shown that when $\varepsilon > 0$, there is a unique rationalizable action: No Attack when $x_i < \bar{x}(\varepsilon, y)$ and Attack when $x_i > \bar{x}(\varepsilon, y)$ where $\bar{x}(\varepsilon, y)$ is a decreasing function of y, approaching 1 as $y \to -\infty$ and 0 as $y \to \infty$. Now consider any $\theta \in (0, 1)$—with multiple equilibria under complete information. We can select y so small that $\bar{x}(\varepsilon, y) > \theta$ for small ε, so that all nearby types (with $x_i \cong \theta$ and $\varepsilon \cong 0$) play No Attack. Similarly, we can select y so large that $\bar{x}(\varepsilon, y) < \theta$ for small ε and hence all nearby types play Attack. The intuition behind this example is that although the private signals are much more informative about θ, the players put more weight on the public signal when they decide what to play.

The second part of our result shows that the alternative selections we were able to achieve in Example 2 are also quite general: Consider any model with a finite type space and *any* rationalizable strategy s in this model. We can always introduce a suitable form of small incomplete information so that for each type t in the original game, $s(t)$ is the unique rationalizable action in an open set of situations close to t. We can then select any rationalizable outcome we want by considering a suitable class of perturbations, and this outcome is robust to further small perturbations. Therefore, selection among rationalizable strategies becomes a matter of selecting among the information structures that look similar to the situation.

To put this another way, multiplicity in the original model comes back in a stronger form when we have only limited information about the players' beliefs (in particular, when we do not have information concerning the entire infinite hierarchy of beliefs). Now, we know that our solution concept would lead to a unique solution in a "more complete" model where our assumptions are relaxed, but we could not know what that outcome would be without knowing in which way the model should be completed. This is a stronger form of multiplicity because it will be true for any non-empty refinement of rationalizability, as the refinement must pick the unique outcome, which changes as we consider different ways to relax our assumptions. This immediately implies that in order for a prediction based on a refinement to be robust to alternative specifications of beliefs, it must be true for all rationalizable strategies, and the robust predictions of refinements are precisely the predictions of rationalizability itself. (We will formally establish this characterization as an immediate application of our theory.)

Formulating the above results for general games inherently requires topological notions on large spaces, and interpretation of such results requires great care. In the next section, we will explain and justify our formulation and describe and interpret our formal results. In Section 3, we introduce the model and preliminary results. We present our results about sensitivity of rationalizable strategies to higher-order beliefs, generic uniqueness, structure of rationalizability, and robustness of predictions of arbitrary refinements in Sections 4, 5, 6, and 7, respectively. We discuss the literature in Section 8. Section 9 concludes. Some proofs are relegated to the appendix.

2. A NON-TECHNICAL EXPOSITION OF OUR FORMULATION AND RESULTS

Because of the nature of our results, our analysis and some of our terminology is technical, but our conclusions have important implications for applied game theory that, through careful interpretation, can be understood without technical jargon. In this section, we carefully describe our formulation and interpret our results in a way that is precise but does not require a heavy mathematical background.

The seed concept of our paper is that common-knowledge assumptions play a crucial role in game-theoretical predictions. The predictions when it is common knowledge that $\theta = \theta_0$ may significantly differ from those when everybody knows that $\theta = \theta_0$, everybody knows this, everybody knows that everybody knows this ... only up to a large but finite order k. Rubinstein (1989) illustrated this with the e-mail game.

Example 3 (E-mail Game, Rubinstein (1989)). In Example 1, assume that $\theta \in \Theta = \{-2/5, 2/5, 6/5\}$. Write $T = \left\{t^{CK}(2/5)\right\}$ for the model in which it is common knowledge that $\theta = 2/5$. Now imagine an incomplete information game in which the players may find it possible that $\theta = -2/5$. Ex ante, players assign probability 1/2 to each of the values $-2/5$ and $2/5$. Player 1 observes the value of θ and automatically sends a message if $\theta = 2/5$. Each player automatically sends a message back whenever he receives one, and each message is lost with probability 1/2. When a message is lost the process automatically stops, and each player is to take one of the actions of Attack or No Attack. This game can be modeled by the type space $\tilde{T} = \{-1, 1, 3, 5, \ldots\} \times \{0, 2, 4, 6, \ldots\}$, where the type t_i is the total number of messages sent or received by player i (except for type $t_1 = -1$ who knows that $\theta = -2/5$), and the common prior p on $\Theta \times \tilde{T}$ where $p(\theta = -2/5, t_1 = -1, t_2 = 0) = 1/2$ and for each integer $m \geq 1$, $p(\theta = 2/5, t_1 = 2m-1, t_2 = 2m-2) = 1/2^{2m}$ and $p(\theta = 2/5, t_1 = 2m-1, t_2 = 2m) = 1/2^{2m+1}$. Here, for $k \geq 1$, type k knows that $\theta = 2/5$, knows that the other player knows $\theta = 2/5$, and so on through k orders. Now, type $t_1 = -1$ knows that $\theta = -2/5$, and hence his unique rationalizable action is No Attack. Type $t_2 = 0$ does not know

θ but puts probability $2/3$ on type $t_1 = -1$, thus believing that player 1 will play No Attack with at least probability $2/3$, so that No Attack is the only best reply and hence the only rationalizable action. Applying this argument inductively for each type k, one concludes that the new incomplete-information game is dominance-solvable, and the unique rationalizable action for all types is No Attack.

If we replace $\theta = -2/5$ with $\theta = 6/5$, we obtain another model, for which Attack is the unique rationalizable action. We consider type space $\check{T} = \{-1, 1, 3, 5, \ldots\} \times \{0, 2, 4, 6, \ldots\}$ and the common prior q on $\Theta \times \check{T}$ where $q\left(\theta = 6/5, t_1 = -1, t_2 = 0\right) = 1/2$ and for each integer $m \geq 1$, $q\left(\theta = 2/5, t_1 = 2m-1, t_2 = 2m-2\right) = 1/2^{2m}$ and $q(\theta = 2/5, t_1 = 2m-1, t_2 = 2m) = 1/2^{2m+1}$. One can easily check that this game is dominance-solvable, and all types play Attack.

In this example, for each action a_i that is rationalizable in the complete-information case and each k, we found another model with a type who knows that $\theta = 2/5$, that each player knows that $\theta = 2/5$, that each player knows that each player knows that $\theta = 2/5$, ..., up to order k, but for this type, a_i is the unique rationalizable action. In this paper, we generalize the construction of this example to arbitrary games, possibly with incomplete information. In an incomplete-information game, a player may not know θ. But each type always has a belief about θ, which we call his first-order belief, has also a belief about θ and the other players' first-order beliefs, which we call his second-order belief, and so on. Generalizing the construction in the above example, for each rationalizable action a_i of each type t_i, and for each k, we construct another model with a type whose first k orders of beliefs are "almost" identical to those of t_i, but for the new type, a_i is the *only* rationalizable action.

What does this imply for economic modeling? There are two distinct classes of situations with incomplete information. In certain problems, there is an ex ante stage during which each party observes a private signal about the payoffs, and the joint distribution of signals and payoffs is commonly known. These problems are naturally modeled using a standard type space. We focus on the alternative class of situations, namely genuine situations of incomplete information, which we believe to be prevalent. In these situations, there is no ex ante stage; each player begins with some first-order beliefs, some second order beliefs, and so on. It has become standard practice to follow Harsanyi (1967) and model these latter problems by introducing a hypothetical ex ante stage, leading to a standard type space. In constructing a type space to model the players' beliefs, a researcher needs to make (implicit) assumptions about aspects of the beliefs that he cannot directly observe in the modeling stage, for standard type spaces exclude the possibility of some interim beliefs which are very close to those implied by the model.

Our results below on how the set of rationalizable outcomes changes as we perturb the interim beliefs are therefore very revealing as to how the conclusions of the modeler are affected by these unverifiable assumptions we make whenever we construct a type space.

Let us briefly return to Example 3 to illustrate how interim beliefs are the primary focus of our analysis even when we apply our ideas to standard type spaces which include an ex ante stage. In type spaces $\tilde{T}$ and $\check{T}$, we consider types k, for k large, to be close to the complete-information type t^{CK} because their interim beliefs are similar. On the other hand, the type spaces $\tilde{T}$ and $\check{T}$ assign only probability $1/2$ to $\theta = 2/5$, which is assigned probability 1 by the complete-information model. Therefore, those who focus on the ex ante perspective, such as Kajii and Morris (1997), would consider these models to be far from the common-knowledge model, as they require the prior probabilities to be similar. Indeed, they obtain a different conclusion: in their framework there is a nearby model that selects the risk-dominant outcome of No Attack, as does $\tilde{T}$, but not one that selects Attack, as does $\check{T}$. In other words, our construction of $\check{T}$ would have inevitably failed if we had insisted on assigning a large prior probability to $\theta = \theta_0$. While their approach is important for applications where we have specific knowledge of an ex ante stage, we focus on genuine incomplete-information situations, where there is no commonly known ex ante stage, and the objective in constructing a model with prior beliefs is to describe the interim beliefs.

Therefore, in our formulation, we directly consider interim belief hierarchies, each of which corresponds to a type in some type space. The set of all such hierarchies is called the universal type space (Mertens and Zamir (1985), Brandenburger and Dekel (1993)). Our notion of which interim belief hierarchies are close to one another, formally described by a topology on the universal type space, will be crucial to our analysis. We have two important motivations for this choice, which serendipitously lead to the same topology. Firstly, we envision a researcher who is restricted to observing only finitely many orders of beliefs. This is especially plausible because common sense suggests that the players themselves will have their beliefs only partially articulated in their own minds. Therefore, hierarchies should be considered close if they agree, or almost agree, at all orders up to order k for large k. Secondly, we would like our topology to capture the usual notion of continuity in standard models. For example, in Example 1, beliefs are continuous functions of (ε, x_1, x_2),[1] and we would like to consider types corresponding to (ε, x_1, x_2) and $(\varepsilon', x_1', x_2')$ close, when (ε, x_1, x_2) and $(\varepsilon', x_1', x_2')$ are close to each other in the usual sense.

[1] Recall that beliefs are probability distributions, and we put the usual weak topology on probability distributions, the topology corresponding to "convergence in distribution".

Mertens and Zamir (1985) have shown that the product topology is the topology described above. Consider a standard model with a compact state space, with any topology such that the beliefs are continuous functions of states. (In Example 1, a state is $(\varepsilon, \theta, x_1, x_2)$.) Assume that there are no two types with the same hierarchy. Then, when we put the product topology on the belief hierarchies, the function that maps the types to the corresponding belief hierarchies is an isomorphism (i.e. it is one-to-one, continuous, and has a continuous inverse). That is, taking limits on belief hierarchies with respect to the product topology is *equivalent* to taking limits in the original space on the types. Since we would like to be able to enlarge models in a manner in which beliefs remain continuous functions of the states, we will then use the product topology in the universal type space.

The product topology also captures the above restriction on the researcher's ability to observe the players' beliefs, as follows. Let T be the set of belief hierarchies generated by the models that he considers possible. Suppose that his observation indicates, for some k, that for each $l \leq k$, an open set of lth-order beliefs are possible (with respect to the weak topology on probability distributions). Then, the set of types that he finds possible are those types in T whose first k orders of beliefs are in these open sets. The product topology relative to T is the smallest topology under which all such sets of types are open. In this topology, concepts such as openness and denseness have specific meanings:

- openness of a set U means that if the actual type is in U and the researcher's observation is sufficiently precise (i.e. k is large and the noise is small), then he would know that actual type is in U;
- that a type t is on the boundary of an open set U means that whenever t is consistent with his observation, the researcher cannot rule out the possibility that by having more precise information he would come to learn that the actual situation is represented by a type in U;
- denseness of a set V means that he could never rule out the possibility that the actual case is represented by a type in V (even if k is very large and the noise is small);
- if U is open and dense, then all types $t \notin U$ are on the boundary of U, that is, no matter how precise his information is, he cannot rule out the possibility that by having more precise information he would come to learn that the actual situation is represented by a type in U.

We consider a finite set of players and a finite set A of actions. Following Carlsson and van Damme, we assume that each action is strictly dominant for some parameter value. We endow the game with the universal type space T^* with the product topology. Under these conditions, we can give a detailed description of the rationalizability correspondence around finite types (types from finite type spaces).

STRUCTURE OF RATIONALIZABILITY. *Given any finite type t and any rationalizable action a for t, there is an open set U^a of types for which a is uniquely rationalizable and t is on the boundary of U^a.*

That is, if t is consistent with the researcher's observation, then, no matter how precise his observation is (i.e. even if he knows arbitrarily many orders of beliefs with arbitrarily small noise), he could not rule out the possibility that, by having more precise information, he would come to learn that a is the unique rationalizable outcome. In particular, when there are multiple rationalizable actions at t, the researcher always finds it possible that with more information he could learn that any particular one is uniquely rationalizable, but he could not know in advance which one. This result immediately leads to a characterization of the situations with multiplicity:

CHARACTERIZATION OF MULTIPLICITY. *A finite type t has multiple rationalizable actions if and only if t is on the boundary of some two open sets U^a and U^b on which two distinct actions a and b are uniquely rationalizable, respectively.*

That is, a finite type has multiple rationalizable actions if and only if it can be thought of idealizing multiple strategically distinct situations with unique rationalizable actions. For example, in Example 1, a type with $\varepsilon > 0$ and $x_i = 1/2$ can be thought of idealizing two situations: (i) x_i is close to $1/2$ but smaller than $1/2$ and (ii) x_i is close to $1/2$ but larger than $1/2$. These two situations are strategically distinct; No Attack is the unique outcome in (i), while Attack is the unique outcome in (ii). This leads to multiplicity at $x_i = 1/2$. Our characterization states that the same picture applies to all cases of multiplicity (under our topology). This can explain why complete-information models tend to a have a large number of rationalizable strategies. Such a model idealizes all of the situations in which private information is small, which can happen in many different information structures. The information structure may have a significant impact on the outcome even when players have small private information.

Since the set of finite types is dense in the universal type space (Mertens and Zamir (1985)), the above structure of rationalizability leads to the following surprising result.

GENERIC UNIQUENESS. *The set $U \subset T^*$ of types with unique rationalizable action is open and dense, and the unique action is locally constant (i.e. each $t \in U$ has a neighborhood in which the unique rationalizable action is the same across all types).*

Under our topology, the interpretation of this statement is that no matter how precise the researcher's observation is, he could not rule out the possibility that, by having more precise information, he would come to learn that there is a unique rationalizable outcome, and would learn what it is.

In applications, we often use small type spaces, such as finite models with a common prior. Suppose that only a subset of models are considered to be possible, and let T be the set of type profiles generated by these models. Assume that T is dense; i.e., for each possible observation the researcher might have, there is a type profile $t \in T$ that is consistent with that observation. For example, the collection of all finite models with a common prior would be one such set (Mertens and Zamir (1985), Lipman (2003)). Then, since U is open and dense in T^*, $U \cap T$ is open and dense in the relative topology on T. Once again, within this smaller set of models, multiplicity occurs only on a nowhere-dense set. (Recall that the complement of an open and dense set is nowhere-dense.) Also, by the isomorphism of Mertens and Zamir (1985), if we enrich a model by allowing sufficiently many types and maintaining the assumptions above, then multiplicity will occur only on a nowhere-dense set of states in the enriched model with respect to the topology on the model (see Section A.4 in the appendix).

There is an immediate application of our results to robustness of the predictions generated by refinements. Imagine that the above researcher subscribes to a particular refinement. For each possible incomplete-information model, represented by a Bayesian game, the researcher can compute a set of possible strategies using the refinement. He wants to make predictions of the form, "for every solution s that satisfies the refinement, $Q(s)$ is true", e.g., "the type with lowest valuation bids zero," or "the bidder with the highest valuation wins the object." Recall that when he wants to make a prediction, the researcher cannot observe the entire hierarchy of beliefs. There are many types from various models that can lead to interim beliefs that are consistent with his limited observation. We want predictions to be robust to these modeling alternatives, in the sense that they will remain true for each of the models consistent with his observation, given that he would have used the refinement in all the alternative models as well. These are the predictions that can be verified by using the limited information. In the coordinated attack example, neither non-trivial prediction is robust in this way:

Example 4. In Examples 1 and 2, take any $\theta \in (0, 1/2)$. Consider a refinement that selects the (Attack,Attack) equilibrium in the complete information game, as Pareto dominance does. Suppose that, based on this refinement and the complete-information model, the researcher predicts that player i will Attack. The researcher cannot rule out the possibility that the actual situation is as described by a type with $x_i \cong \theta$ and $\varepsilon \cong 0$ in Example 1. For this specification No Attack is the only rationalizable action, and his refinement must assign the action No Attack to this alternative type. Hence, he cannot make this prediction in this alternative specification, using his refinement. Therefore, his prediction of Attack is not robust. The prediction that there will be no attack—as predicted by risk-dominance—is not robust, either, because the researcher cannot rule out that the actual situation is as described by a type with $x_i \cong \theta$, $\varepsilon \cong 0$, and y very large in Example 2, when Attack is the only rationalizable action.

More generally, by our result, given any rationalizable strategy in a finite type space, the researcher cannot rule out the possibility that under an alternative specification the strategy would be uniquely rationalizable—and thus be the unique outcome of his refinement. This leads to a general characterization: *a prediction of a refinement is robust if and only if it is true for all rationalizable strategies in the model.* This characterization suggests that without making any common-knowledge assumption, one cannot make any prediction stronger than what is implied already by rationalizability, no matter how strong a refinement one uses. Of course, a researcher may be willing to make common-knowledge restrictions, or very strong assumptions on possible information structures. In that case, he may be able to make sharper predictions by subscribing to a stronger solution concept. Even then, our characterization will be useful: by considering all rationalizable strategies, the researcher can find out which of his predictions would remain valid if he relaxed these assumptions.

A complementary approach to ours is to study "strategic topologies" under which, by definition, nearby types have similar strategic behavior (see Monderer and Samet (1989) and Dekel, Fudenberg, and Morris (2006)). Certain types which might naturally appear to converge do not converge strategically in the sense that their strategic behavior would not converge.[2] By discovering which types converge under the strategic topology, one then hopes to learn what precision of observation of types is required for accurate predictions of behavior. When there are more converging sequences in the natural topology that do not converge in strategic topology, there will be fewer accurate predictions one can make under the restrictions assumed by the natural topology. In this paper, conversely, we fix a natural topology that captures a reasonable restriction on a researcher's ability to observe players' beliefs. Analyzing the properties of strategic behavior with respect to this topology, we determine which predictions the researcher can make under that restriction. This general approach was promoted by Rubinstein (1989).

[2] In Example 1, when $\varepsilon' > 0$ and $\bar{\theta} \neq 1/2$, $(\varepsilon = 0, \bar{\theta}, \bar{\theta})$ is strategically distinct from any type $(\varepsilon', \bar{\theta}, \bar{\theta})$. As $\varepsilon \to 0$, the types with $x_i = \bar{\theta}$ would not converge to the common-knowledge case in strategic topologies. For another example, Dekel, Fudenberg, and Morris (2006) show that the set of finite types is dense in their strategic topology as it is in the product topology, but the set of finite types with common prior is not dense in their strategic topology as the behavior they exhibit cannot be close to some forms of betting behavior that non-CPA types would permit.

3. MODEL

We consider a finite set of players $N = \{1, 2, \ldots, n\}$. There is a possibly unknown payoff-relevant parameter $\theta \in \Theta^*$ where Θ^* is a compact (and hence complete and separable) metric space. Each player i has a finite action space A_i and utility function $u_i : \Theta^* \times A \to \mathbb{R}$, where $A = \prod_i A_i$.[3] We consider the set of games that differ in their specifications of the belief structure on θ, i.e. their type spaces. By a *model*, we therefore mean a pair $(\Theta \times T, \kappa)$ where $\Theta \subseteq \Theta^*$ and $T = T_1 \times \cdots \times T_n$ is a type space associated with beliefs $\kappa_{t_i} \in \Delta(\Theta \times T_{-i})$ for each $t_i \in T_i$.

Given any type t_i in a model $(\Theta \times T, \kappa)$, we can compute the belief of t_i on Θ^* by first extending[4] κ_{t_i} to $\Theta^* \times T_{-i}$ and setting

$$t_i^1 = \operatorname{marg}_{\Theta^*} \kappa_{t_i},$$

which is called the *first-order belief of* t_i. We can compute the *second-order belief of* t_i, i.e. his belief about $(\theta, t_1^1, \ldots, t_n^1)$, by setting

$$t_i^2(F) = \kappa_{t_i}\left(\left\{(\theta, t_{-i}) \mid \left(\theta, t_i^1, t_{-i}^1\right) \in F\right\}\right)$$

for each measurable $F \subseteq \Theta^* \times \Delta(\Theta^*)^n$. We can compute an entire hierarchy of beliefs $\left(t_i^1, t_i^2, \ldots, t_i^k, \ldots\right)$ by proceeding in this way. Let us write $h_i(t_i)$ for the resulting hierarchy. We say that a type space T *does not have redundant types* if each h_i is one-to-one; i.e., distinct types have distinct belief hierarchies.

The universal type space is the type space that consists of such belief hierarchies, as in the formulation of Brandenburger and Dekel (1993), which we will follow. In this type space, a type of a player i is an infinite hierarchy of beliefs

$$t_i = \left(t_i^1, t_i^2, \ldots\right)$$

where $t_i^1 \in \Delta(\Theta^*)$ is a probability distribution on Θ^*, representing the beliefs of i about θ, $t_i^2 \in \Delta(\Theta^* \times \Delta(\Theta^*)^n)$ is a probability distribution for $(\theta, t_1^1, t_2^1, \ldots, , t_n^1)$, representing the beliefs of i about θ and the other players' first-order beliefs, and so on. The set of all belief hierarchies for which it is common knowledge that the beliefs are coherent (i.e., each player knows his beliefs and his beliefs at different orders are consistent with

[3] Our notation is standard. Specifically, given any list $Y_1, \ldots, Y_n$ of sets, write $Y = \prod_i Y_i$, $Y_{-i} = \prod_{j \neq i} Y_j$, $y_{-i} = (y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n) \in Y_{-i}$, and $(y_i, y_{-i}) = (y_1, \ldots, y_{i-1}, y_i, y_{i+1}, \ldots, y_n)$. Likewise, for any family of functions $f_j : Y_j \to Z_j$, we define $f_{-i} : Y_{-i} \to Z_{-i}$ by $f_{-i}(y_{-i}) = (f_j(y_j))_{j \neq i}$. Given any metric space (Y, d), we write $\Delta(Y)$ for the space of probability distributions on Y, endowed with Borel σ-algebra and the weak topology. We use the product σ-algebra in product spaces. We also write $\operatorname{supp}(\pi)$ for the support of a probability distribution π, $\operatorname{marg}_Y \pi$ for the marginal of π on Y, and proj_Y for the projection mapping to Y.

[4] The extension simply puts probability zero on the remaining set $(\Theta^* \backslash \Theta) \times T_{-i}$.

each other) is denoted by T_i^*. $T^* = T_1^* \times \cdots \times T_n^*$ denotes the set of all type profiles $t = (t_1, \ldots, t_n)$, and $T_{-i}^* = \prod_{j \neq i} T_j^*$ is the set of profiles of types t_{-i} for players other than i. Each T_i^* is endowed with the product topology, so that a sequence of types $t_{i,m}$ converges to a type t_i, denoted by $t_{i,m} \to t_i$, if and only if $t_{i,m}^k \to t_i^k$ for each k. A sequence of type profiles $t(m) = (t_{1,m}, \ldots, t_{n,m})$ converges to t iff $t_{i,m} \to t_i$ for each i. For each type t_i, let $\kappa_{t_i} \in \Delta\left(\Theta^* \times T_{-i}^*\right)$ be the unique probability distribution that represents the beliefs of t_i about (θ, t_{-i}). Mertens and Zamir (1985) have shown that the mapping $t_i \mapsto \kappa_{t_i}$ is an isomorphism. That is, it is one-to-one, and $\kappa_{t_{i,m}} \to \kappa_{t_i}$ if and only if $t_{i,m} \to t_i$. Finally, models without redundant types can be identified with certain subsets T of universal type space T^*, and when it does not lead to confusion, we refer to such subsets of T^* as models.[5]

In our formulation, it is common knowledge that the payoffs are given by a fixed continuous function of parameters. This assumption is without loss of generality because we could take a parameter to be simply the function that maps action profiles to payoff profiles. For example, we can take $\Theta^* = \Theta_1^* \times \cdots \times \Theta_n^*$ where $\Theta_i^* = [0,1]^A$ for each i, and let $u_i(\theta, a) = \theta_i(a)$ for each (i, a, θ). This model allows all possible payoff functions, and here θ is simply an index for the profile of payoff functions. This model clearly satisfies the following richness assumption, which is also made by Carlsson and van Damme (1993).

Assumption 1 (Richness Assumption). *For each i and each a_i, there exists $\theta^{a_i} \in \Theta^*$ such that*

$$u_i\left(\theta^{a_i}, a_i, a_{-i}\right) > u_i\left(\theta^{a_i}, a_i', a_{-i}\right) \qquad \left(\forall a_i' \neq a_i, \forall a_{-i}\right).$$

That is, the space of possible payoff structures is rich enough so that each action can be strictly dominant for some parameter value. When there are no a priori restrictions on the domain of payoff structures and the game is static, Assumption 1 is automatically satisfied. In a dynamic game, one needs to introduce trembles and use a reduced form to satisfy this assumption.[6] It would be important future work to extend our results to

[5] For any model $(\Theta \times T, \kappa)$, the set $h(T) \subseteq T^*$ is *belief-closed*, i.e., for each $h_i(t_i) \in h_i(T_i)$, we have $\kappa_{h_i(t_i)}\left(\Theta^* \times h_{-i}(T_{-i})\right) = 1$. Conversely, any belief-closed subset $T \subseteq T^*$ corresponds to a model.

[6] Given any finite extensive-form game, two strategies are equivalent if they lead to the same outcome for each profile of the other players' strategies. Let A_i contain one strategy from each equivalence class. Introduce trembles, in the sense that after each move of a player, Nature moves and changes the player's action to some other action with a small probability. The other players can see the realized action, but not the intended one. Then, by varying the payoffs at the terminal nodes, we can make each $a_i \in A_i$ dominant at some payoff function.

dynamic games without introducing trembles, for many traditional refinements play a role mainly in dynamic games.

A *strategy* of a player i with respect to T_i is any function $s_i : T_i \to A_i$.[7] When the domain of a strategy is omitted, it is understood to be T_i^*. For each $i \in N$ and for each belief $\pi \in \Delta(\Theta \times A_{-i})$, we write $BR_i(\pi)$ for the set of actions $a_i \in A_i$ that maximize the expected value of $u_i(\theta, a_i, a_{-i})$ under the probability distribution π.

Interim Correlated Rationalizability. For each i and t_i, set $S_i^0[t_i] = A_i$, and define sets $S_i^k[t_i]$ for $k > 0$ iteratively, by letting $a_i \in S_i^k[t_i]$ if and only if $a_i \in BR_i\left(\mathrm{marg}_{\Theta^* \times A_{-i}} \pi\right)$ for some $\pi \in \Delta\left(\Theta^* \times T_{-i}^* \times A_{-i}\right)$ such that $\mathrm{marg}_{\Theta^* \times T_{-i}^*} \pi = \kappa_{t_i}$ and $\pi\left(a_{-i} \in S_{-i}^{k-1}[t_{-i}]\right) = 1$. That is, a_i is a best response to a belief of t_i that puts positive probability only to the actions that survive the elimination in round $k-1$. We write $S_{-i}^{k-1}[t_{-i}] = \prod_{j \neq i} S_j^{k-1}[t_j]$ and $S^k[t] = S_1^k[t_1] \times \cdots \times S_n^k[t_n]$. The set of all rationalizable actions for player i (with type t_i) is

$$S_i^\infty[t_i] = \bigcap_{k=0}^{\infty} S_i^k[t_i].$$

A strategy $s_i : T_i^* \to A_i$ is said to be rationalizable iff $s_i(t_i) \in S_i^\infty[t_i]$ for each t_i.

For complete-information games, rationalizability has been introduced by Bernheim (1984) and Pearce (1984). There are several notions of rationalizability for incomplete-information games. Interim correlated rationalizability (Battigalli (2003), Battigalli and Siniscalchi (2003) and Dekel, Fudenberg, and Morris (2003)) is the weakest among these notions, as shown by Dekel, Fudenberg, and Morris (2003). Using such a weak notion of rationalizability strengthens our results; they will remain valid under any stronger notion of rationalizability. In the definition of rationalizability, we did not refer to a model because it only depends on the belief hierarchy of the type, as stated in the following lemma. This result shows that we can analyze the properties of rationalizability by focusing on the universal type space.[8]

Lemma 1 (Dekel, Fudenberg, Morris (2003)). *Given any model $(\Theta \times T, \kappa)$ and any $t \in T$, $S_i^\infty[t_i] = S_i^\infty[h_i(t_i)]$. Moreover, for any $t_i, \tilde{t}_i \in T_i^*$ with $t_i^l = t_i^l$ for all $l \leq k$, we have $S_i^k[t_i] = S_i^k[\tilde{t}_i]$.*

We conclude this section by introducing some familiar concepts.

[7] We do not restrict the strategies to be measurable; measurability is not needed in the present interim framework (cf. Simon (2003)).

[8] Ely and Peski (2006) show that if one wants to define a different concept, independent rationalizability, and there are redundant types, then one needs to consider a larger type space.

DEFINITION 1 (Finite Types, Models). A model $(\Theta \times T, \kappa)$ is said to be *finite* iff $|\Theta \times T| < \infty$. Let $\hat{T}$ be the set of all profiles of belief hierarchies corresponding to a finite model i.e., $\hat{t}_i \in \hat{T}_i$ iff $\hat{t}_i \in h_i(T_i)$ for some finite model $(\Theta \times T, \kappa)$. Members of $\hat{T}$ are referred to as *finite types*.

DEFINITION 2 (Dominance-Solvability). A model $T \subseteq T^*$ is said to be *dominance solvable* if and only if $|S^\infty[t]| = 1$ for each $t \in T$.

DEFINITION 3 (Common Prior). A finite model $(\Theta \times T, \kappa)$ is said to *admit a common prior (with full support)* if and only if there exists a probability distribution $p \in \Delta(\Theta \times T)$ with support $\Theta \times T$ and such that $\kappa_{t_i} = p(\cdot|t_i)$ for each $t_i \in T_i$. Such a model is also denoted by $(\Theta \times T, p)$. The set of all profiles of belief hierarchies that come from a model with a common prior is denoted by

$$T_i^{CPA} = \{h_i(t_i)\,|t_i \in T_i \text{ for some finite model } (\Theta \times T, \kappa) \text{ with a common prior}\}.$$

4. SENSITIVITY TO HIGHER-ORDER BELIEFS

In this section we show that, when there are multiple rationalizable actions at a finite type t_i, all rationalizable strategies are highly sensitive to the perturbations of beliefs at t_i. In particular, given any rationalizable action a_i, we can perturb the beliefs of t_i slightly and obtain a type for which a_i is uniquely rationalizable as stated in the following result.

PROPOSITION 1. *Under Assumption 1, for any $\hat{t} \in \hat{T}$, and any $a \in S^\infty[\hat{t}]$, there exists a sequence of finite, dominance-solvable models $T^m \subset \hat{T}$ with type profiles $\tilde{t}(m) \in T^m$, such that $\tilde{t}(m) \to \hat{t}$ as $m \to \infty$ and $S^\infty[\tilde{t}(m)] = \{a\}$ for each m.*

That is, given any k, we can perturb the first k orders of beliefs slightly and vary the higher-order beliefs arbitrarily to find a type $\tilde{t}_i$ for which a_i is the unique rationalizable action, i.e., $S_i^\infty[\tilde{t}_i] = \{a_i\}$. Hence, even if we have a very good idea about what the first k orders of beliefs are, by varying the higher-order beliefs we can still make any rationalizable action uniquely rationalizable. Moreover, we can find such types in finite type spaces that are dominance-solvable.

Our proof of this result is the heart of our technical contribution. It is notationally involved and is relegated to the appendix, but the key ideas are simple, as we will summarize below. As a preliminary step, we can perturb the beliefs of $\hat{t}_i$ slightly to make a rationalizable action a_i "strictly rationalizable", in the sense that it survives iterated elimination of actions that are never a *strict* best reply. This is possible because we can always break the ties for best reply in favor of a desired action by allowing the type to put slightly higher probability to the payoff function at which that action is dominant. Moreover, we can introduce the perturbations in such a way that the perturbed type space remains finite. We can then focus on the case that the action at hand is "strictly" rationalizable.

Our main step shows that if an action a_i survives the first k rounds of iterated elimination of actions that are never a strict best reply, we can change the beliefs at order $k+1$ and higher in such a way that a_i is the *only* action that survives the first $k+1$ rounds of elimination of strictly dominated actions. Indeed, we can do this in such a way that the resulting type space is finite and dominance-solvable in $k+1$ rounds, i.e., $S^{k+1}[t]$ is singleton for each type profile t in the new type space. We can then conclude that if a_i is "strictly" rationalizable for a finite type $\bar{t}_i$, then for each k we can find a finite dominance-solvable model with a type $\tilde{t}_i$ such that the first k orders of beliefs coincide with those of $\bar{t}_i$ (i.e. $\tilde{t}_i^l = \bar{t}_i^l$ for each $l \leq k$) and a_i is the unique rationalizable action at $\tilde{t}_i$ (i.e. $S_i^\infty\left[\tilde{t}_i\right] = \{a_i\}$).

Our proof of the main step generalizes the construction in e-mail game of Section 2 to arbitrary incomplete information games with finite type spaces. To see the main idea, consider a two-player game. By Assumption 1, each action a_j is dominant for some type $t_j[a_j]$ who is certain that $\theta = \theta^{a_j}$, and $S_j^1[t_j[a_j]] = \{a_j\}$. This is the $k=0$ case. For the $k=1$ case, now consider a type t_i and an action a_i that survives the first round of the elimination of actions that are never a strict best reply. Then, a_i is the unique best reply to a belief p of t_i on pairs (θ, a_j). Consider the type $\tilde{t}_i$ that puts probability $p(\theta, a_j)$ on $(\theta, t_j[a_j])$ for each pair (θ, a_j). Now, at the second round of elimination, $\tilde{t}_i$ can entertain only one belief about the actions of each type he finds possible: $t_j[a_j]$ plays a_j with probability 1. He thus assigns the same probability to (θ, a_j) as to $(\theta, t_j[a_j])$, which is also exactly $p(\theta, a_j)$. But a_i is the unique best reply to this belief, and therefore $S_i^2\left[\tilde{t}_i\right] = \{a_j\}$. Moreover, $\tilde{t}_i^1 = t_i^1$, i.e., the beliefs of $\tilde{t}_i$ and t_i on Θ are identical, because to any θ, type $\tilde{t}_i$ assigns probability $\sum_{a_j} p(\theta, t_j[a_j]) = \sum_{a_j} p(\theta, a_j)$, and the latter sum is the probability that t_i assigns to θ, by definition of p. This completes the argument for $k=1$; we are able to repeat this argument inductively on k.

5. GENERICITY OF UNIQUENESS

In this section, we show that the set

$$U = \{t \in T^* | \, |S^\infty[t]| = 1\}$$

of type profiles with unique rationalizable action profiles is open and dense. We first introduce the mathematical notions and existing results that are necessary to present our result.

Definition 4 (Genericity). The *closure* of a set $T \subseteq T^*$, denoted by $\overline{T}$, is the smallest closed set that contains T. A set T is *dense* (in T^*) iff $\overline{T} = T^*$, i.e., for each $t \in T^*$, there exists a sequence of type profiles $t(m) \in T$ such that $t(m) \to t$. A set T is said to be *nowhere-dense* iff the interior of $\overline{T}$ is empty, i.e., $\overline{T}$ does not contain any open set.

An open and dense set $T \subseteq T^*$ is large in the sense that we can approximate each $\tilde{t} \in T^* \backslash T$ by type profiles $t \in T$, and we cannot approximate any $t \in T$ by type profiles $\tilde{t} \in T^* \backslash T$. In this case, $T^* \backslash T$ is simply the boundary of T, denoted by ∂T. Being open and dense is a strong topological notion of genericity. Topological notions of genericity may differ widely from measure-theoretical notions of genericity, which are about how commonly an event occurs under a measure (see Oxtoby (1980) for more on the relationship between these notions). They also depend on the topology. That is to say, our genericity result may not be true under other topologies or under measure-theoretical notions of genericity. Hence, one should be careful in interpreting it as saying that there are few types with multiple rationalizable actions, as the precise notion of "few" is crucial. The precise meaning of our particular notion of genericity was discussed in Section 2.

Mertens and Zamir (1985) show that finite types are dense (i.e. $\overline{\hat{T}} = T^*$). Lipman (2003) further shows that, in finite models, the common-prior assumption does not put any restriction on finite-order beliefs other than full support[9] (see also Feinberg (2000)), proving the following result.

Lemma 2 (Mertens and Zamir (1985) and Lipman (2003)). *$\hat{T}$ and T^{CPA} are dense in the universal type space, i.e., $\overline{T^{CPA}} = \overline{\hat{T}} = T^*$.*

That is, the finite models with common prior, which are predominantly used in economic modeling, lead to a dense set of type profiles. We will also use the following properties of rationalizability.

[9] Consider any finite model $(\Theta \times T, \kappa)$ where each κ_{t_i} has full support. For any k, Lipman constructs another finite model $\left(\Theta \times \tilde{T}, p\right)$ with common prior such that for each type $t_i \in T_i$ there is a type $\tau(t_i) \in \tilde{T}_i$ such that the first k orders of beliefs under t_i and $\tau(t_i)$ are identical. The main idea is that we can replace the belief differences in lower-order beliefs that are due to lack of a common prior with belief differences that come from informational differences. The higher-order beliefs in these two situations will differ, but Lipman shows that we can move these differences to arbitrarily high orders.

LEMMA 3 (Dekel, Fudenberg, and Morris (2006)). *S^∞ is non-empty and upper-semicontinuous. That is, each $t \in T^*$ has a neighborhood η with $S^\infty[t'] \subseteq S^\infty[t]$ for each $t' \in \eta$.*

Armed with these tools and our Proposition 1, we can now state and prove our main result in this section:

PROPOSITION 2. *Under Assumption 1, the set*

$$U = \{t \in T^* | \, |S^\infty[t]| = 1\}$$

of type profiles with unique rationalizable action profiles is open and dense. Moreover, the unique rationalizable outcome is locally constant on U, i.e., each $t \in U$ has an open neighborhood η on which S^∞ is constant.

Proof. To show that U is dense, first observe that, by Proposition 1, for any $\hat{t} \in \hat{T}$, there exists a sequence $\tilde{t}(m) \to \hat{t}$ with $S^\infty\left[\tilde{t}(m)\right] = \{a\}$ for some $a \in S^\infty\left[\hat{t}\right]$. By definition, $\tilde{t}(m) \in U$ for each m. Hence, $\bar{U} \supseteq \hat{T}$. But $\overline{\hat{T}} = T^*$ by Lemma 2. Therefore, $\bar{U} \supseteq \overline{\hat{T}} = T^*$, showing that U is dense. On the other hand, Lemma 3 immediately implies that U is also open, as we show now. By upper-semicontinuity of S^∞, each $t \in U$ has a neighborhood η with $S^\infty[t'] \subseteq S^\infty[t]$ for each $t' \in \eta$. Since $S^\infty[t'] \neq \varnothing$ and $S^\infty[t]$ is singleton, this implies that $S^\infty[t'] = S^\infty[t]$ for each $t' \in \eta$, showing that $\eta \subset U$. Therefore, U is open. This also establishes the last statement in the proposition. □

By Proposition 2, we can partition the universal type space into an open and dense set U and its nowhere-dense boundary $T^* \backslash U$. On U, each type has a unique rationalizable action, and every rationalizable strategy is continuous. On the boundary, each type profile has multiple rationalizable action profiles. Assumption 1 is not superfluous. For example, a complete-information game can be modeled with $|\Theta^*| = 1$, when T^* consists of a single common-knowledge type profile. When the original game is not dominance-solvable, U is empty.

Harsanyi and Selten (1988, p. 341) seek "rational criteria for selecting *one* equilibrium point as the solution of any non-cooperative game". This is in line with the more general perspective that a complete description of the environment should lead to a unique outcome. Proposition 2 shows that there is a specific sense in which common knowledge of rationality itself leads to a unique solution in "generic" situations. To reach such a unique solution, one does not need to assume any additional rationality criteria or include anything beyond the payoffs and the payoff-related information structure in the description of the environment.

One may wonder if the genericity result above applies to smaller type spaces of interest, such as the space of finite types and space of finite types consistent with the common-prior assumption. The next result shows that the same genericity result is true for any dense type space, including these spaces.

Corollary 1. *Under Assumption 1, for any dense model $T \subseteq T^*$, the set $U \cap T$ is dense and open with respect to the relative topology on T. In particular, $U \cap (T^{CPA})$ is dense and open with respect to the relative topology on T^{CPA}.*

Proof. Since U is open and dense and T is dense, $U \cap T$ is dense. Since U is open, $U \cap T$ is open with respect to the relative topology on T, by definition. □

6. STRUCTURE OF RATIONALIZABILITY

We will now turn to the major question of why a particular type may have multiple rationalizable actions. Focusing on finite models, we will provide an answer to this question, and uncover a striking structure of the rationalizability correspondence on the universal type space. Towards this end, we first show that, for each finite model and for each rationalizable strategy profile s_T in this model, we can perturb the beliefs and find a new dominance-solvable model such that s_T is the unique rationalizable strategy profile for the perturbed types.

Proposition 3. *Let $T \subseteq \hat{T}$ be any finite model and $s_T : T \to A$ be any rationalizable strategy profile. Then, under Assumption 1, there exist a sequence of finite dominance-solvable models $T^{s_T,m}$ and a sequence of one-to-one mappings $\tau(\cdot, s_T, m) : T \to T^{s_T,m}$ such that, for each $t \in T$,*

1. *$S^\infty[\tau(t, s_T, m)] = \{s_T(t)\}$, and*
2. *$\tau(t, s_T, m) \to t$ as $m \to \infty$.*

Proof. By Proposition 1, for each $t \in T$ and m, there exists a finite, dominance-solvable model $T^{t,s_T,m}$ with $\tau(t, s_T, m) \in T^{t,s_T,m}$ as in the proposition. Define the finite model $T^{s_T,m}$ by

$$T_i^{s_T,m} = \bigcup_{t \in T} T_i^{t,s_T,m}.$$

Since $\tau(t, s_T, m) \to t$ for each $t \in T$ and T is finite, there exists $\bar{m}$ such that, for any distinct t, t' and any $m > \bar{m}$, we have $\tau(t, s_T, m) \neq \tau(t', s_T, m)$. Hence, $\tau(\cdot, s_T, m)$ is one-to-one for $m > \bar{m}$. (Consider only $m > \bar{m}$.) □

This result extends the result of Carlsson and van Damme to arbitrary finite models (for extension to infinite models, see Yildiz (2005)): we can always perturb a model by introducing a small noise in players' perceptions of the payoffs in such a way that the new model is dominance-solvable. Moreover, since U is open, the perturbed model will remain dominance-solvable when we introduce new small perturbations. But unlike in Carlsson and van Damme, we can select any rationalizable strategy we want by introducing a suitable form of incomplete information. The dominance-solvable model $T^{s_T,m}$ in our proof is a "collage" of disparate submodels and need not admit a common prior. One might think that our selection of arbitrary rationalizable strategies in nearby dominance-solvable models stems from lack of a common prior. Building on Lipman (2003), the next result shows that this is not the case.

Proposition 4. *Let $T \subseteq \hat{T}$ be any finite model and $s_T : T \to A$ be any rationalizable strategy profile, with $s_T(t) \in S^\infty[t]$ for each $t \in T$. Then, under Assumption 1, there exist sequences of finite models $\tilde{T}^{s_T,m}$ with common prior and one-to-one mappings $\tilde{\tau}(\cdot, s_T, m) : T \to \tilde{T}^{s_T,m}$ such that*

1. *$S^\infty[\tilde{\tau}(t, s_T, m)] = \{s_T(t)\}$, and*
2. *$\tilde{\tau}(t, s_T, m) \to t$ as $m \to \infty$ for each $t \in T$.*

Since the proof of this result is somewhat involved, we present it in the appendix. The key idea is simple: since the finite types with common prior are dense, for each $\tau(t, s_T, m)$, we can find a nearby type profile that comes from a finite model with a common prior. Consider the "collage" of these models, which admits a common prior. By introducing a further perturbation to this collage model, we obtain a model $\tilde{T}^{s_T,m}$ where the common prior has full support. Since the rationalizable actions are robust to small perturbations in dominance-solvable models (by the last statement in Proposition 1), the perturbed type $\tilde{\tau}(t, s_T, m)$ in the latter model has the same rationalizable action as the perturbed type $\tau(t, s_T, m)$ in the dominance-solvable model, which is $s_T(t)$.

Propositions 3 and 4 uncover a striking structure of rationalizability on the set $\hat{T}$ of finite types. This structure remains intact when one imposes the common-prior assumption (i.e. on T^{CPA}). One can divide $\hat{T}$ into finitely many open sets

$$U^a = \left\{\hat{t} \in \hat{T} | S^\infty\left[\hat{t}\right] = \{a\}\right\} \qquad (a \in A)$$

and their boundaries $\partial U^a \equiv \overline{U^a} \backslash U^a$, where $\overline{U^a}$ is the closure of U^a, all with respect to the relative topology on $\hat{T}$. The open sets form a partition of an open, dense set $U \cap \hat{T}$, while their boundaries cover the boundary of $U \cap \hat{T}$, i.e., $\hat{T} \backslash U = \bigcup_{a \in A} \partial U^a$, which is a nowhere-dense set with respect to the relative topology. On each open set U^a, a is the unique rationalizable action profile. Since S^{∞} is upper-semicontinuous, $a \in S^{\infty}\left[\hat{t}\right]$ for each $\hat{t} \in \partial U^{a}$. Hence, at any $\hat{t} \in \partial U^{a} \cap \partial U^{a'}$ with distinct a and a', both a and a' are rationalizable. Here, there are multiple rationalizable actions a and a' because $\hat{t}$ can be thought of idealization of two strategically distinct relaxed assumptions, under which a and a' are unique solutions respectively, and the set of rationalizable actions reflects this fact. Propositions 3 and 4 show that the converse is also true:

$$\hat{t} \in \bigcap_{a \in S^{\infty}\left[\hat{t}\right]} \overline{U^{a}} \qquad \left(\forall \hat{t} \in \hat{T}\right).$$

That is, the set $S^{\infty}\left[\hat{t}\right]$ tells us precisely which actions could be uniquely rationalizable when we slightly relax the assumptions of $\hat{t}$ using various information structures. When $\hat{t}$ has multiple rationalizable actions, it cannot be in the interior of any of these sets. This leads to the following result.

Corollary 2. *Under Assumption 1, for any $\hat{t} \in \hat{T}$ with $\left|S^{\infty}\left[\hat{t}\right]\right| > 1$,*

$$\hat{t} \in \bigcap_{a \in S^{\infty}\left[\hat{t}\right]} \partial U^{a}.$$

In particular, for any $\hat{t} \in \hat{T}$, $\left|S^{\infty}\left[\hat{t}\right]\right| > 1$ if and only if there exist distinct $a, b \in A$ such that $\hat{t} \in (\partial U^{a}) \cap \left(\partial U^{b}\right)$.

That is, whenever there are multiple rationalizable actions at $\hat{t}$, $\hat{t}$ embodies an idealization of multiple possible relaxed assumptions with distinct strategic implications. For each rationalizable action a, there is such a relaxed assumption, which leads to a as the unique solution. Rationalizability is then a generically unique and locally constant solution concept that yields multiple solutions at, and only at, the boundaries where the concept changes its prescribed behavior.

Proposition 4 also provides a new perspective on refining rationalizability. It implies that a finite model summarizes various dominance-solvable situations by abstracting away from the details that would have mattered mostly for computing the beliefs at very high orders. By specifying these details appropriately, any rationalizable strategy could have been made uniquely rationalizable. But then, refining rationalizability is tantamount to ruling out some of these nearby models as the true model. Hence, under the assumptions of our paper, selection of a refinement is tied to the assumptions on mutual beliefs *about payoff parameters*—more so than to the assumptions typically made in refinements about mutual beliefs *about strategies*.

7. Robustness to higher-order beliefs

In this section we will formalize our notion of robustness to higher-order beliefs and characterize the set of robust predictions for arbitrary refinements of rationalizability, including equilibrium refinements. We find that the only robust predictions are those that are true for all rationalizable strategies, i.e. those that could have been made without the refinement. We start by formally presenting some basic definitions.

Definition 5. A *solution concept* is any mapping Σ that maps each model $M = (\Theta \times T, \kappa)$ to a set $\Sigma(M)$ of distributions σ over strategy profiles with respect to T. An *equilibrium refinement* is any solution concept Σ such that for each M and $\sigma \in \Sigma(M)$, σ is a Bayesian Nash equilibrium of M. A *refinement of rationalizability* is any solution concept Σ such that for each M and $\sigma \in \Sigma(M)$, $\sigma(s) > 0$ implies that $s(t) \in S^{\infty}[t]$ for each $t \in T$.

We speak of distributions over strategy profiles in order to allow correlation between the strategies as in correlated equilibrium. Clearly, any equilibrium refinement is a refinement of rationalizability. Moreover, rationalizability, S^{∞}, is a solution concept, yielding all possible distributions on the strategy profiles with $s(t) \in S^{\infty}[t]$ for each $t \in T$ at each $M = (\Theta \times T, \kappa)$.

Definition 6. Given a solution concept Σ and a model $M = (\Theta \times T, \kappa)$, by a *prediction* of (Σ, M), we mean any formula Q with free variable $s : T \to A$ such that $Q(s)$ is true for each $s \in \mathrm{supp}(\sigma)$ and each $\sigma \in \Sigma(M)$.

A prediction can be about the behavior of a particular type. In an auction, for example, a prediction could be "the type with lowest valuation bids zero." A prediction can also be about a relation between the behavior of different types, e.g., "a player's bid is an increasing function of his valuation."

We envision a researcher who subscribes to an equilibrium refinement Σ and can observe players' beliefs up to an arbitrary but finite order k with some noise. We focus on the case that k is large and the noise is small. The researcher may also want to restrict the set of models by fiat. For example, it is customary in economics literature to assume that there is a common prior, and he may want to impose the common-prior assumption on possible models, ignoring the models without a common prior altogether. He may also want to focus on small models, such as finite type spaces. Taking these considerations into account, we will assume that the researcher considers only the finite models with common prior, so that he does not want to check whether his predictions would remain valid in models with infinite types or with non-common priors. These are clearly unrealistically generous assumptions, but we will show that, despite this, the researcher cannot make very strong predictions.

In our formulation the noise will be small in the sense that, for each $l \leq k$, the lth-order beliefs $\tilde{t}_i^l$ that the researcher finds possible converge to the observed beliefs t_i^l in the sense of "convergence in distribution", i.e., in the weak topology. (Recall that t_i^l is a probability distribution.) To do this, we consider an arbitrary metric d on finite-order beliefs that metrizes the weak topology. Given the observed or estimated beliefs, t_i^l, $l \leq k$, the researcher finds the set of beliefs $\tilde{t}_i^l$ with $d\left(\tilde{t}_i^l, t_i^l\right) \leq \epsilon$ for all $l \leq k$ possible (or cannot reject them at a particular level of confidence) for some $\epsilon > 0$, where ϵ is meant to measure the precision of the researcher's observations. Again, we focus on the limit $\epsilon \to 0$ and $k \to \infty$.

Definition 7. Given a model $M = (\Theta \times T, \kappa)$, a pair $(\tilde{T}, \tau)$ of a model $\tilde{T} \subset T^*$ and a mapping $\tau : T \to \tilde{T}$ is said to be an (ϵ, k)-*perturbation* of M iff (i) $\tilde{T}$ is finite and has a common prior, and (ii) $\tau : T \to \tilde{T}$ is such that for each $t \in T$ and $l \leq k$, we have $d\left(\tilde{t}_i^l, \hat{t}_i^l\right) \leq \epsilon$ where $\tilde{t} = \tau(t)$ and $\hat{t} = h(t)$ is the belief hierarchy of t.

The definition of an (ϵ, k)-perturbation requires that whenever our researcher believes that a type profile t in T may describe the actual situation, he cannot rule out that the type profile $\tau(t)$ in $\tilde{T}$ describes the situation. That is, the researcher cannot reject the perturbation without rejecting the original model. A perturbation may result from relaxing the assumption that a certain fact is common knowledge, instead assuming that it is approximately mutually known only up to kth order, and perhaps making some other assumption about the higher-order beliefs. Reflecting such a relaxation, the perturbed model will then have more type profiles. This is the case in the e-mail game of Section 2. As we discussed, type k (in $\tilde{T}$ or in $\check{T}$) agrees with the common knowledge type $t_i^{CK}(2/5)$ up to order k. Therefore, both $(\tilde{T}, \tau)$ and $\left(\check{T}, \tau\right)$ are $(0, k)$-perturbations of the complete information game T for any mapping τ whose values are both at least k.

Our robustness condition will require that the prediction remains valid for all perturbations defined as above, for some ϵ and k. The motivation is clear. Imagine a researcher analyzing a model M, knowing that when he is asked to validate that his model applies to a particular situation, he will be able to observe only the first k orders of beliefs, with some noise. He knows that, if he can verify that M is consistent with his observation, $\tilde{T}$ will also be consistent with his observation. If a prediction of his model M does not remain true for the perturbation $\tilde{T}$, then he cannot verify that his prediction applies to the situation. Therefore, the researcher would like to focus on the predictions of M that are robust to alternative specifications, such as $\tilde{T}$, given that he was going to apply his solution concept in those alternative specifications as well. We therefore define robustness as follows.

Definition 8. A prediction Q of (Σ, M) is said to be (ϵ, k)-*robust (to higher-order beliefs)* iff for each (ϵ, k)-perturbation $(\tilde{T}, \tau)$ of M, for each $\sigma \in \Sigma(\tilde{T})$, for each $s \in \text{supp}(\sigma)$, $Q(s \circ \tau)$ is true. Prediction Q is said to be *robust (to higher-order beliefs)* iff it is (ϵ, k)-robust for some $\epsilon > 0$ and $k < \infty$.

That is, a prediction is said to be (ϵ, k)-robust if it remains true in models where the first k orders of beliefs remain close to the original beliefs according to the perturbation mapping and we apply the same solution concept throughout. We could weaken our robustness requirement by requiring $Q(s \circ \tau)$ to be true only if s is the "unique" solution in the perturbed model, i.e., $\text{supp}(\sigma(\tau(t))) = \{s(t)\}$ at each $t \in T$. It will be clear that our results would remain valid under this substantially weaker requirement. Returning again to the e-mail game example, the prediction of no attack for the complete information game $T = \left\{t^{CK}(2/5)\right\}$ is not robust under any equilibrium refinement Σ because for each k, $\left(\check{T}, \tau\right)$ with $\tau\left(t^{CK}(2/5)\right) > (k, k)$ is a $(0, k)$-perturbation of T, and for the unique member σ of $\Sigma\left(\check{T}\right)$, $\sigma\left(\tau\left(t^{CK}(2/5)\right)\right)$ assigns probability 1 to (Attack,Attack). Similarly, the prediction of Attack is not robust.

Both of the non-robustness results in this example are special cases of the upcoming proposition. Characterizing the robust predictions of any refinement, it states that no refinement can make robust predictions that are any more powerful than the predictions that are generated by rationalizability.

Proposition 5. *Under Assumption 1, for any equilibrium refinement Σ that is nonempty on finite models with a common prior and any finite model M, a prediction Q of (Σ, M) is robust if and only if Q is a prediction of (S^∞, M).*

Proof. Consider a finite model $M = (\Theta \times T, \kappa)$. We first show that any robust prediction Q of (Σ, M) is a prediction of (S^∞, M). Any such Q is an (ϵ, k)-robust prediction of (Σ, M) for some $\epsilon > 0$ and $k < \infty$. Take any $s : T \to A$ with $s(t) \in S^\infty[t]$ for each $t \in T$. By Proposition 4, there exist a finite model $\tilde{T} \subset T^{CPA}$ with common prior and a mapping $\tilde{\tau} : h(T) \to \tilde{T}$ such that for each $t \in T$ and $\tilde{t} = \tilde{\tau}(h(t))$ we have $d\left(\tilde{t}_i^l, \hat{t}_i^l\right) \leq \epsilon$, where $\hat{t} = h(t)$, and $S^\infty\left[\tilde{t}\right] = \{s(t)\}$. Since Σ is a refinement of rationalizability, this implies that $\Sigma(\tilde{T}) = \{\tilde{s}\}$ where $\tilde{s} \circ \tilde{\tau} \circ h = s$. Since $(\tilde{T}, \tilde{\tau} \circ h)$ is an (ϵ, k)-perturbation of M

and Q is (ϵ, k)-robust, this further implies that $Q(s) = Q(\tilde{s} \circ \tilde{\tau} \circ h)$ is true. Therefore, Q is a prediction of (S^∞, M).

For the converse, take any prediction Q of (S^∞, M). By Lemma 3, there exist $\epsilon > 0$ and $k < \infty$ such that $S^\infty\left[\tilde{t}\right] \subseteq S^\infty[t]$ whenever $t \in T$ and $d\left(\tilde{t}_i^l, \hat{t}_i^l\right)$ for all $l \leq k$, where $\hat{t}_i = h_i(t_i)$. Hence, for any (ϵ, k)-perturbation $(\tilde{T}, \tau)$ of M and any $s \in \text{supp}(\sigma)$ with $\sigma \in \Sigma(\tilde{T})$, we have $s(\tau(t)) \in S^\infty[t]$ for each $t \in T$. Since Q is a prediction of (S^∞, M), this shows that $Q(s \circ \tau)$ is true. Therefore, Q is an (ϵ, k)-robust prediction of (Σ, M). □

In our characterization, we only assume that Σ is non-empty on the finite models with a common prior. Since it is customary to prove such an existence result whenever a refinement is proposed, this assumption allows most equilibrium refinements. Moreover, our perturbation considers only the finite models with a common prior. Hence, the non-robustness implied by our characterization is not due to large models or failure of the common-prior assumption.

We must, however, emphasize that our characterization does rely on the interim perspective we take throughout the paper, as our formulation of predictions and robustness are based on that perspective. *If* there is an ex ante stage with a common prior, then the researchers will usually be interested in results that hold with high probability under that prior, and they can also use the prior to form a belief about the actual type in the interim stage. In that case, they can make stronger predictions in the interim stage by using a stronger solution concept than rationalizability, as illustrated in the following example.

Example 5. In Example 3, assume that the researcher assigns prior probability $\tilde{\pi}$ to $\tilde{T}$, $\check{\pi}$ to $\check{T}$, and $1 - \tilde{\pi} - \check{\pi} > 0$ to the complete-information case, T. In the grander model the researcher considers, the probability of $t^{CK}(2/5)$ is $1 - \tilde{\pi} - \pi$; the probability of (θ, t_1, t_2) from $\tilde{T}$ is $\tilde{\pi} p(\theta, t_1, t_2)$, and the probability of (θ, t_1, t_2) from $\tilde{T}$ is $\check{\pi} q(\theta, t_1, t_2)$. Notice that p and q are geometrically decreasing with t_1 and t_2. If a researcher is interested only in behavior that occurs with probability higher than some P, then he can ignore all the types t_i from $\tilde{T}$ or $\check{T}$ that are higher than some integer $K(P)$, the types that lead to the non-robustness results in the e-mail game. The researcher can do more. Suppose that he observes the first k orders of beliefs for some large k. If the actual case is described by a type $k' < k$ from either model $\tilde{T}$ or $\check{T}$, then the researcher will know the type of the player, precisely. If the actual type is some $k' > k$ or we are in the common-knowledge case, then he will observe kth-order mutual knowledge of $\theta = 2/5$. There are types from all three models consistent with this observation, but he will assign a probability that is nearly 1 to the common knowledge case (by Bayes rule). Then, his refinement for the complete information case will allow him to make sharper predictions (regardless of whether he subscribes to risk-dominance or Pareto-dominance).

8. LITERATURE REVIEW

Sensitivity to Higher-order Beliefs and Global Games. In the context of his e-mail game (Example 3), Rubinstein (1989) illustrated that the predictions of Pareto-dominant equilibrium may be highly sensitive to the specification of higher-order beliefs. Subsequently, Carlsson and van Damme (1993) showed for two-player, two action, supermodular games that when one introduces small incomplete-information as in Example 1, the risk-dominant action becomes uniquely rationalizable. They then argued that we should select the risk-dominant equilibrium in the original game. This led to many well-known applications, such as Morris and Shin (1998). Frankel, Morris, and Pauzner (2003) extended the uniqueness result of Carlsson and van Damme to all supermodular games with complete information but illustrated in an example that the selected outcome may depend on the noise structure. In the context of Example 2, Morris and Shin (2000) also illustrated that the selected outcome may depend on the prior when the noise in the private signal is not negligible with respect to the prior. In this paper, we generalize both uniqueness and noise-dependence results in a strong way: (i) we can make *any game* dominance-solvable, by introducing a suitable form of small incomplete information, but (ii) by varying the form of incomplete information, we can select *any* rationalizable strategy in the original game, weakening the selection argument.

Robust Equilibrium. Kajii and Morris (1997) introduced a notion of robustness of a given equilibrium of a given complete-information game to incomplete information, as follows. They define an equilibrium of a complete information game to be robust iff, in any incomplete-information model M with common prior assigning high probability to the event that the payoffs are as described in the complete-information game and everybody knows his payoffs, M will have an equilibrium in which most of the types will play according to the original equilibrium. This concept of robustness rules out incomplete information games that involve large changes in prior but may lead to interim beliefs that are similar to the actual situation. For example, in Example 2, they require that $y = x_i$. They also exclude the e-mail game, if the probability of $\theta \neq 2/5$ is not close to zero. As we have shown, however, at the interim stage, a researcher could not know that probability without having knowledge of the entire infinite hierarchy of beliefs. Then, the key difference between our notions of perturbation is that they focus on small changes to *prior* beliefs, without regard to the size of changes to *interim* beliefs, while our focus is the reverse. Their approach is appropriate when there is an ex-ante stage along with well-understood inference rules and we know the prior to some degree. As we discussed in Section 2, however, in genuine incomplete-information situations, the type spaces and ex-ante stage are just tools for modeling interim beliefs. In that case, it is appropriate to consider types with similar interim beliefs, even if they come from models that assign small prior probability to the actual situation.

Payoff-irrelevant or Epistemic Types. Brandenburger and Dekel (1987) have shown that given a distribution on rationalizable strategy profiles of a given complete-information game, we can enrich the type space by adding payoff-irrelevant types and find an *equilibrium* in the new game that yields the same distribution on the strategy profiles of the original game.[10] That is, in order for a prediction to be robust with respect to the entire set of equilibria *without any refinements*, it must be true for all rationalizable strategies. We show that *for any refinement* of rationalizability (or equilibrium), when we allow other payoff-relevant types from alternative models that lead to similar interim beliefs, in order for a prediction gained by the refinement to be robust, it must be true for all rationalizable strategies.

In this paper we drop all common-knowledge restrictions. Of course, some may want to make explicit common-knowledge restrictions on players' payoffs and beliefs. Battigalli and Siniscalchi (2003) introduce a notion of Δ-rationalizability, which corresponds to common-knowledge of such assumptions and rationality. Under such restrictions, it seems that a similar analysis to ours would show that the predictions of any refinement that remain valid with only partial knowledge of interim beliefs will be equivalent to that of Δ-rationalizability.

Robustness of Equilibrium Refinements. Our approach on robustness is closest to that of Fudenberg, Kreps, and Levine (1988) and Dekel and Fudenberg (1990). Using types in the spirit of Assumption 1, Fudenberg, Kreps, and Levine (1988) have shown that any equilibrium of any complete-information game can be made strict by perturbing the payoffs, showing that one cannot obtain any more predictions than those true for all equilibria by considering refinements that do not eliminate any strict equilibrium. Our result covers refinements that do eliminate some strict equilibria, such as the popular risk-dominance, and compares them to the larger set of all rationalizable strategies for arbitrary information structures.

In the same vein, Dekel and Fudenberg (1990) analyze the robustness of predictions based on iterated elimination of weakly dominated strategies when one allows payoff uncertainty as in this paper. Maintaining the common prior assumption, they show that with uncertainty about players' beliefs at all orders, the robust predictions gained from this procedure for a complete-information game is equivalent to those of iterated

[10] This result has been extended to incomplete-information games by Battigalli and Siniscalchi (2003) and Dekel, Fudenberg, and Morris (2003).

strict dominance after one round of eliminating weakly dominated strategies.[11] Dropping the common-prior assumption, they also show that even if we know that each player's prior puts high probability to original payoffs, we could not rule out the possibility that a strategy that survives the latter elimination process is a strict equilibrium action. Hence, under this limited knowledge, the "robust" predictions of a refinement that does not eliminate any strict equilibrium are no more than the predictions of the latter solution concept.

Role of common-prior assumption. Our results in this paper remain the same whether the common-prior assumption holds or fails. Unlike here, the common-prior assumption plays a central role in many existing robustness results. For example, if we drop the common-prior assumption in the Kajii and Morris definition, then existence of a robust equilibrium implies that $n-1$ players have dominant strategies (see Weinstein and Yildiz (2004) and Oyama and Tercieux (2005)). This is because without a common prior, their restrictions on prior beliefs have no implications for interim beliefs beyond second order. Likewise, if we impose the common-prior assumption in Brandenburger and Dekel, all types in their type spaces must play a correlated equilibrium strategy (by Aumann (1987)). Similarly, the construction of Dekel and Fudenberg for their second result crucially relies on their departure from the common-prior assumption (as in Brandenburger and Dekel).

Sensitivity to higher-order beliefs in applications. Following the critique of Wilson (1987), a sizeable literature has established that some central findings in economics, such as the full surplus extraction property of Cremer and Mclean (1988) in mechanism design (Neeman (2004) and Heifetz and Neeman (2006)) and the Coase conjecture in bargaining (Feinberg and Skrzypacz (2005)), crucially rely on the assumptions on the second-order beliefs and higher. In this paper, we show that this sensitivity is a general phenomenon.

9. CONCLUSION

We can sum up our main results by describing the following intuitive picture, partitioning the universal type space (with a rich set of payoff functions) according to the set of rationalizable outcomes. "Most" of the space is taken up by a family of open sets, one for each action, namely the set of types for which this is the unique rationalizable

[11] Borgers (1994) shows that the latter solution concept charcterizes the strategies that are consistent with almost-common knowledge of players not playing weakly dominated strategies. Here, almost-common knowledge is in the sense of common p-belief by Monderer and Samet (1989). Monderer and Samet show that an equilibrium remains an approximate equilibrium (similar to the robust equilibrium of Kajii and Morris) if there is common p-belief of the original game for high p, but we cannot check this condition without knowledge of the infinite hierarchy of beliefs.

action. These sets and their boundaries indeed constitute the entire space, and at each boundary we have multiple rationalizable actions, namely the actions corresponding to the neighboring open sets. This picture has two important implications for game theory, particularly for refinements. First, for an open and dense set of types, there is a unique rationalizable action, implying that all refinements must agree on a unique solution. This suggests that the ubiquity of large multiplicity of rationalizable outcomes in present models may be due to the special structure of these games, rather than being an inherent property of rationalizable behavior. Second, when a researcher is restricted to observing only finite-order beliefs with some noise, then for every rationalizable action a, he must find it possible that if he had a more precise but still limited knowledge of beliefs, he could have learned that a is the unique rationalizable action for the actual situation, when his refinement would necessarily select a as the unique solution. In order for a prediction, based on his refinement, to remain valid in the actual situation, it must remain true when the solution is a. Consequently, a prediction remains valid under such alternative specifications if and only if it could be deduced from rationalizability alone. Therefore, without a common-knowledge restriction in the full, infinite-order sense, refinements do not lead to any new robust predictions.

Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208; j-weinstein@kellogg.northwestern.edu; http://www1.kellogg.northwestern.edu/facdir/facpage.asp?sid=1299

and

MIT Economics Department, 50 Memorial Dr., Cambridge, MA 02142; myildiz@mit.edu; http://econ-www.mit.edu/faculty/myildiz/index.htm.

Appendix A. proofs and further results

In this appendix, we will present the proofs and some other details that are omitted in the main text. The following existing result will be very useful in our proofs; it shows that any model without redundant types is isomorphic to a subset of this space.

Lemma 4 (Mertens and Zamir (1985) and Brandenburger and Dekel (1993)). *Let $(\Theta \times T, \kappa)$ be any model, endowed with any topology, such that $\Theta \times T$ is complete and separable and κ_{t_i} is a continuous function of t_i. Then, h is continuous. Moreover, if $(\Theta \times T, \kappa)$ does not have redundant types and $\Theta \times T$ is compact, then h is an isomorphism (i.e., both h and its inverse are continuous).*

A.1. *The Details of Example* 2

Given any $\varepsilon > 0$ and y, conditional on x_i,

$$(\theta, x_j) \sim N\left(\left(\begin{array}{c} rx_i + (1-r)\,y \\ rx_i + (1-r)\,y \end{array}\right), \left[\begin{array}{cc} \varepsilon^2 r & \varepsilon^2 r \\ \varepsilon^2 r & \varepsilon^2 (r+1) \end{array}\right]\right), \tag{A.1}$$

where $r = 1/\left(1+\sqrt{2\pi}\varepsilon\right)$. As $\varepsilon \to 0$, r converges to 1, and the above distribution converges to the point mass at (x_i, x_i), according to which it is common knowledge that $\theta = x_j = x_i$. Lemma 4 implies that, as $\varepsilon \to 0$, the belief hierarchy of type x_i converges to the belief hierarchy that states that it is common knowledge that $\theta = x_j = x_i$. By (A.1), the cutoff value $\bar{x}(\varepsilon, y)$ is the unique solution to the indifference equation

$$rx + (1-r)\,y = \Phi\left(\frac{(1-r)(x-y)}{\varepsilon\sqrt{1+r}}\right), \tag{A.2}$$

where Φ is the cumulative distribution function of the standard normal distribution. When ε is small, $r \cong 1 - \sqrt{2\pi}\varepsilon$, and hence

$$\bar{x}(\varepsilon, y) \cong \Phi\left(\sqrt{\pi}\left(\bar{x}(\varepsilon, y) - y\right)\right).$$

As we increase y to ∞, $\lim_{\varepsilon\to 0} \bar{x}(\varepsilon, y)$ decreases to 0, and as we decrease y to $-\infty$, $\lim_{\varepsilon\to 0} \bar{x}(\varepsilon, y)$ increases to 1.

A.2. *Proof of Proposition* 1

In our proof, we will refer to a strict version of rationalizability, denoted by W^∞, where we eliminate an action if it is not a strict best reply to any belief. We define W^∞, by setting $W_i^0[t_i] = A_i$ and letting $a_i \in W_i^k[t_i]$ if and only if $BR_i\left(\mathrm{marg}_{\Theta^*\times A_{-i}}\pi\right) = \{a_i\}$ for some $\pi \in \Delta\left(\Theta^* \times T_{-i}^* \times A_{-i}\right)$ such that $\mathrm{marg}_{\Theta^*\times T_{-i}^*}\pi = \kappa_{t_i}$ and $\pi\left(a_{-i} \in W_{-i}^{k-1}[t_{-i}]\right) = 1$. The set of all *strictly rationalizable* actions for t_i is

$$W_i^\infty[t_i] = \bigcap_{k=0}^{\infty} W_i^k[t_i].$$

The set W^∞ has the following familiar fixed point property.

Lemma 5. *Given any model $(\Theta \times T, \kappa)$, consider any family $V_i[t_i] \subseteq A_i$, $t_i \in T_i$, $i \in N$, such that each $a_i \in V_i[t_i]$ is a strict best reply to a belief $\pi \in \Delta(\Theta \times T_{-i} \times A_{-i})$ of t_i with $\mathrm{marg}_{\Theta\times T_{-i}}\pi = \kappa_{t_i}$ and $\pi(a_{-i} \in V_{-i}[t_{-i}]) = 1$. Then, $V_i[t_i] \subseteq W_i^\infty[t_i]$ for each t_i.*

The first, and a preliminary, step of our proof is to show that by perturbing the beliefs in a finite model slightly, we can make all the rationalizable actions strictly rationalizable:

Lemma 6. *Under Assumption 1, for any finite model $T \subseteq \hat{T}$, there exists a sequence of finite models T^m, $m = 1, 2, \ldots$, with bijections $\tau(\cdot, m)$ from the set of pairs (t, a) with $t \in T$ and $a \in S^\infty[t]$ to T^m, such that (i) $a \in W^\infty[\tau(t, a, m)]$ for each (t, a, m), and (ii) $\tau(t, a, m) \to t$ as $m \to \infty$ for each (t, a).*

Proof. Let Θ' be the finite set of all parameter values that some type $t_j \in T_j$ assigns positive probability. For each $a_i \in S_i^{\infty}[t_i]$, there exists a belief $\pi^{t_i,a_i} \in \Delta(\Theta' \times T_{-i} \times A_{-i})$ with $a_i \in BR_i\left(\mathrm{marg}_{\Theta' \times A_{-i}} \pi^{t_i,a_i}\right)$, $\pi^{t_i,a_i}\left(a_{-i} \in S_{-i}^{\infty}[t_{-i}]\right) = 1$, and $\mathrm{marg}_{\Theta \times T_{-i}} \pi^{t_i,a_i} = \kappa_{t_i}$. Now, for each $\varepsilon \in [0,1]$, consider the model $(\Theta \times T^{\varepsilon}, \kappa)$, where $\Theta = \Theta' \cup \{\theta^{a_i} | a_i \in A_i, i \in N\}$ and each T_i^{ε} consists of types $\bar{\tau}_i(t_i, a_i, \varepsilon)$, $i \in N$, $t_i \in T_i$, and $a_i \in S_i^{\infty}[t_i]$, defined by

$$\kappa_{\bar{\tau}_i(t_i,a_i,\varepsilon)} = \varepsilon \delta_{\left(\theta^{a_i}, \bar{\tau}_{-i}\left(\tilde{t}_{-i}, \tilde{a}_{-i}, \varepsilon\right)\right)} + (1-\varepsilon)\, \pi^{t_i,a_i} \circ \hat{\tau}_{-i,\varepsilon}^{-1}$$

where δ_x denote the probability distribution that puts probability 1 on $\{x\}$, $\bar{\tau}_{-i}\left(\tilde{t}_{-i}, \tilde{a}_{-i}, \varepsilon\right)$ is some fixed type profile in T_{-i}^{ε}, and $\hat{\tau}_{-i,\varepsilon} : (\theta, t_{-i}, a_{-i}) \mapsto (\theta, \bar{\tau}_{-i}(t_{-i}, a_{-i}, \varepsilon))$. For each $\bar{\tau}_i(t_i, a_i, \varepsilon)$, define the belief

$$\tilde{\pi} = \kappa_{\tau_i(t_i,a_i,\varepsilon)} \circ \gamma^{-1} \in \Delta\left(\Theta \times T_{-i}^{\varepsilon} \times A_{-i}\right)$$

where $\gamma : (\theta, \bar{\tau}_{-i}(t_{-i}, a_{-i}, \varepsilon)) \mapsto (\theta, \bar{\tau}_{-i}(t_{-i}, a_{-i}, \varepsilon), a_{-i})$, so that $\bar{\tau}_i(t_i, a_i, \varepsilon)$ believes that a_{-i} is played at each $(\theta, \bar{\tau}_{-i}(t_{-i}, a_{-i}, 0))$. Then, by construction,

$$\mathrm{marg}_{\Theta \times A_{-i}} \tilde{\pi} = \varepsilon \delta_{(\theta^{a_i}, \tilde{a}_{-i})} + (1-\varepsilon)\, \mathrm{marg}_{\Theta \times A_{-i}} \pi^{t_i,a_i}.$$

The belief of $\bar{\tau}_i(t_i, a_i, \varepsilon)$ about $\Theta \times A_{-i}$ is a mixture. With probability $(1-\varepsilon)$, $\bar{\tau}_i(t_i, a_i, m)$ faces the same uncertainty as t_i does when t_i holds the belief π^{t_i,a_i}, in which case a_i is a best reply. With probability ε, the equality $\theta = \theta^{a_i}$ holds, in which case a_i is the unique best reply. Then, when $\varepsilon > 0$, a_i is a strict best reply, i.e., $BR_i\left(\mathrm{marg}_{\Theta \times A_{-i}} \tilde{\pi}\right) = \{a_i\}$. Hence, by Lemma 5, $a_i \in W_i^{\infty}[\bar{\tau}_i(t_i, a_i, \varepsilon)]$ for each $\tau_i(t_i, a_i, \varepsilon)$ and $\varepsilon > 0$.

We will now show that $h_i(\bar{\tau}_i(t_i, a_i, \varepsilon)) \to t_i$ as $\varepsilon \to 0$. By construction, each probability distribution $\kappa_{\bar{\tau}_i(t_i,a_i,\varepsilon)}$ is continuous in (t_i, a_i, ε). Hence, by Lemma 4, $h_i(\bar{\tau}_i(t_i, a_i, \varepsilon)) \to h_i(\bar{\tau}_i(t_i, a_i, 0))$ as $\varepsilon \to 0$. Therefore, it suffices to show that $h_i(\bar{\tau}_i(t_i, a_i, 0)) = t_i$ for each t_i and i. To do this, we first note that the first-order beliefs are equal:

$$h_i^1(\bar{\tau}_i(t_i, a_i, 0)) \equiv \mathrm{marg}_{\Theta^*} \kappa_{\bar{\tau}_i(t_i,a_i,0)} = \mathrm{marg}_{\Theta^*} \pi^{t_i,a_i} = \mathrm{marg}_{\Theta^*} \kappa_{t_i} \equiv t_i^1,$$

where the second and third equalities are by definitions of $\hat{\tau}_{-i}$ and π^{t_i,a_i}, respectively. Now fix some $k > 1$, and let L be the set of all belief profiles of players other that i at order $k-1$. Towards an induction, assume that $h_j^{k-1}(\bar{\tau}_j(t_j, a_j, 0)) = t_j^{k-1}$ for each $\bar{\tau}_j(t_j, a_j, 0)$ and j. Then, $\mathrm{proj}_{\Theta^* \times L} \circ \hat{\tau}_{-i,m} = \mathrm{proj}_{\Theta^* \times L}$, and hence

$$\mathrm{marg}_{\Theta^* \times L} \pi^{t_i,a_i} \circ \hat{\tau}_{-i,0} = \mathrm{marg}_{\Theta^* \times L} \pi^{t_i,a_i}.$$

Therefore,

$$h_i^k(\bar{\tau}_i(t_i, a_i, 0)) = \delta_{h_i^{k-1}(\bar{\tau}_i(t_i,a_i,0))} \times \mathrm{marg}_{\Theta^* \times L} \pi^{t_i,a_i} \circ \hat{\tau}_{-i,0}^{-1} = \delta_{h_i^{k-1}(\bar{\tau}_i(t_i,a_i,0))} \times \mathrm{marg}_{\Theta^* \times L} \pi^{t_i,a_i} = t_i^k,$$

showing that $h_i^k(\bar{\tau}_i(t_i, a_i, 0)) = t_i^k$ for each k.

Moreover, there exist $\bar{\varepsilon} > 0$ such that $h_i(\bar{\tau}_i(t_i, a_i, \varepsilon)) \neq h_i(\bar{\tau}_i(t_i', a_i', \varepsilon))$ whenever $(t_i, a_i) \neq (t_i', a_i')$ and $0 < \varepsilon \leq \bar{\varepsilon}$. [For any $a_i \neq a_i'$, by definition, $\theta^{a_i} \neq \theta^{a_i'}$, rendering $h_i(\bar{\tau}_i(t_i, a_i, \varepsilon)) \neq h_i(\bar{\tau}_i(t_i, a_i', \varepsilon))$ when $\varepsilon > 0$. For any $t_i \neq t_i'$, since $h_i(\bar{\tau}_i(t_i, a_i, \varepsilon)) \to t_i$ and $h_i(\bar{\tau}_i(t_i', a_i', \varepsilon)) \to$

t_i', there exists $\bar{\varepsilon}$ such that $h_i\left(\bar{\tau}_i\left(t_i, a_i, \varepsilon\right)\right) \neq h_i\left(\bar{\tau}_i\left(t_i', a_i', \varepsilon\right)\right)$ whenever $\varepsilon \leq \bar{\varepsilon}$.] We pick $T^m = h\left(T^{\bar{\varepsilon}/m}\right)$ and $\tau_i\left(t_i, a_i, m\right) = h_i\left(\bar{\tau}_i\left(t_i, a_i, \bar{\varepsilon}/m\right)\right)$ everywhere. □

Lemma 6 states that, given any rationalizable action a_i, by perturbing the beliefs of the type slightly, we can make a_i *strictly* rationalizable. Our next result, which is the main step in our proof, establishes that by perturbing the beliefs further, we can make this strictly rationalizable action, a_i, uniquely rationalizable, leading to the desired conclusion.

Lemma 7. *Under Assumption 1, for each i, k, for each $\hat{t}_i \in \hat{T}_i$, and for each $a_i \in W_i^k\left[\hat{t}_i\right]$, there exists $\tilde{t}_i$ such that (i) $\tilde{t}_i^{k'} = \hat{t}_i^{k'}$ for each $k' \leq k$, (ii)*

$$S_i^{k+1}\left[\tilde{t}_i\right] = \left\{a_i\right\},$$

and (iii) $\tilde{t}_i \in T_i^{\tilde{t}_i}$ for some finite model $T^{\tilde{t}_i} = T_1^{\tilde{t}_i} \times \cdots \times T_1^{\tilde{t}_i}$ with $\left|S^{k+1}\left[t\right]\right| = 1$ for each $t \in T^{\tilde{t}_i}$. Therefore, for any $a_i \in W_i^{\infty}\left[\hat{t}_i\right]$, there exists a sequence of finite, dominance-solvable models T^m, $m = 1, 2, \ldots$, with types $t_{i,m} \in T_i^m$, such that $S_i^{\infty}\left[t_{i,m}\right] = \left\{a_i\right\}$ and $t_{i,m} \to \hat{t}_i$ as $m \to \infty$.

Proof. For $k = 0$, let $\tilde{t}$ be the type profile according to which it is common knowledge that each j assigns probability 1 to $\left\{\theta = \theta^{a_j}\right\}$, where θ^{a_j} is as defined in Assumption 1. By Assumption 1, $S_i^1\left[\tilde{t}_i\right] = \left\{a_i\right\}$, and it is vacuously true that $\tilde{t}_i^l = \hat{t}_i^l$ for each $l \leq k$. Clearly, the type space $\left\{\tilde{t}\right\}$ is belief-closed.

Now fix any $k > 0$ and any i. Write each t_{-i} as $t_{-i} = (l, h)$ where $l = \left(t_{-i}^1, t_{-i}^2, \ldots, t_{-i}^{k-1}\right)$ and $h = \left(t_{-i}^k, t_{-i}^{k+1}, \ldots\right)$ are the lower and higher-order beliefs, respectively. Let $L = \left\{l | \exists h : (l, h) \in T_{-i}^*\right\}$. The inductive hypothesis is that for each finite $t_{-i} = (l, h)$ and each $a_{-i} \in W_{-i}^{k-1}\left[t_{-i}\right]$, there exists finite $\tilde{t}_{-i}\left[a_{-i}\right] = \left(l, \tilde{h}\left[l, a_{-i}\right]\right) \in T_{-i}^{\tilde{t}_{-i}\left[a_{-i}\right]}$ such that

$$S_{-i}^k\left[\tilde{t}_{-i}\left[a_{-i}\right]\right] = \left\{a_{-i}\right\}, \tag{IH}$$

and $T^{\tilde{t}_{-i}\left[a_{-i}\right]} = T_1^{\tilde{t}_{-i}\left[a_{-i}\right]} \times \cdots \times T_n^{\tilde{t}_{-i}\left[a_{-i}\right]}$ is a finite model with $\left|S^k\left[t\right]\right| = 1$ for each $t \in T^{\tilde{t}_{-i}\left[a_{-i}\right]}$. Take any $a_i \in W_i^k\left[\hat{t}_i\right]$. We will construct a type $\tilde{t}_i$ as in the lemma. By definition, $BR_i\left(\text{marg}_{\Theta^* \times A_{-i}}\pi\right) = \left\{a_i\right\}$ for some $\pi \in \Delta\left(\Theta^* \times T_{-i}^* \times A_{-i}\right)$ such that $\text{marg}_{\Theta^* \times T_{-i}^*}\pi = \kappa_{t_i}$ and $\pi\left(a_{-i} \in W_{-i}^{k-1}\left[t_{-i}\right]\right) = 1$. Using the inductive hypothesis, define mapping $\mu : \text{supp}\left(\text{marg}_{\Theta^* \times L \times A_{-i}}\pi\right) \to \Theta^* \times T_{-i}^*$, by

$$\mu : \left(\theta, l, a_{-i}\right) \mapsto \left(\theta, l, \tilde{h}\left[l, a_{-i}\right]\right), \tag{A.3}$$

where type $\tilde{t}_{-i}\left[a_{-i}\right] = \left(l, \tilde{h}\left[l, a_{-i}\right]\right)$ is as in (IH). Define $\tilde{t}_i$ by

$$\kappa_{\tilde{t}_i} \equiv \left(\text{marg}_{\Theta^* \times L \times A_{-i}}\pi\right) \circ \mu^{-1} = \pi \circ \text{proj}_{\Theta^* \times L \times A_{-i}}^{-1} \circ \mu^{-1}, \tag{A.4}$$

where proj_X denotes the projection mapping to X. By construction of μ, the first k orders of beliefs (about (θ, l)) are identical under t_i and $\tilde{t}_i$:

$$\begin{aligned}
\mathrm{marg}_{\Theta^* \times L} \kappa_{\tilde{t}_i} &= \pi \circ \mathrm{proj}^{-1}_{\Theta^* \times L \times A_{-i}} \circ \mu^{-1} \circ \mathrm{proj}^{-1}_{\Theta^* \times L} \\
&= \pi \circ \mathrm{proj}^{-1}_{\Theta^* \times L} = \left(\pi \circ \mathrm{proj}^{-1}_{\Theta^* \times T^*_{-i}}\right) \circ \mathrm{proj}^{-1}_{\Theta^* \times L} = \mathrm{marg}_{\Theta \times L} \kappa_{t_i},
\end{aligned}$$

where the first equality is by (A.4), the second equality is by (A.3), which implies that $\mathrm{proj}_{\Theta^* \times L} \circ \mu \circ \mathrm{proj}_{\Theta^* \times L \times A_{-i}} = \mathrm{proj}_{\Theta^* \times L}$, and the last equality is by definition of π. Moreover, by (IH), each $(\theta, t_{-i}) \in \mathrm{supp}\big(\kappa_{\tilde{t}_i}\big)$, which is of the form $\left(\theta, l, \tilde{h}\left[l, a_{-i}\right]\right)$, has a unique action $a_{-i} \in S^{k-1}_{-i}\left[\tilde{t}_{-i}\left[a_{-i}\right]\right]$. Thus, there exists a unique $\tilde{\pi} \in \Delta\left(\Theta^* \times T^*_{-i} \times A_{-i}\right)$ such that $\mathrm{marg}_{\Theta^* \times T^*_{-i}} \pi = \kappa_{\tilde{t}_i}$ and $\pi\left(a_{-i} \in S^{k-1}_{-i}\left[t_{-i}\right]\right) = 1$. This belief is $\tilde{\pi} = \kappa_{\tilde{t}_i} \circ \gamma^{-1} = \pi \circ \mathrm{proj}^{-1}_{\Theta^* \times L \times A_{-i}} \circ \mu^{-1} \circ \gamma^{-1}$ where $\gamma : \left(\theta, l, \tilde{h}\left[l, a_{-i}\right]\right) \mapsto \left(\theta, l, \tilde{h}\left[l, a_{-i}\right], a_{-i}\right)$. By construction,

$$\begin{aligned}
\mathrm{marg}_{\Theta^* \times L \times A_{-i}} \tilde{\pi} &= \kappa_{\tilde{t}_i} \circ \gamma^{-1} \circ \mathrm{proj}^{-1}_{\Theta^* \times L \times A_{-i}} \\
&= \pi \circ \mathrm{proj}^{-1}_{\Theta^* \times L \times A_{-i}} \circ \mu^{-1} \circ \gamma^{-1} \circ \mathrm{proj}^{-1}_{\Theta^* \times L \times A_{-i}} = \mathrm{marg}_{\Theta^* \times L \times A_{-i}} \pi.
\end{aligned}$$

[By definition of μ and γ, $\mathrm{proj}_{\Theta \times L \times A_{-i}} \circ \gamma \circ \mu$ is the identity mapping, yielding the last equality.] Thus,

$$\mathrm{marg}_{\Theta^* \times A_{-i}} \tilde{\pi} = \mathrm{marg}_{\Theta^* \times A_{-i}} \pi.$$

But a_i is the only best reply to this belief:

$$BR_i\left(\mathrm{marg}_{\Theta^* \times A_{-i}} \tilde{\pi}\right) = BR_i\left(\mathrm{marg}_{\Theta^* \times A_{-i}} \pi\right) = \left\{a_i\right\}.$$

Therefore, $S^{k+1}_i\left[\tilde{t}_i\right] = \{a_i\}$.

Now, we will define $T^{\tilde{t}_i}$ as in the lemma. Define

$$\begin{aligned}
T^{\tilde{t}_i}_i &= \left\{\tilde{t}_i\right\} \cup \left(\bigcup_{(\theta, t_{-i}[a_{-i}]) \in\, \mathrm{supp}\left(\kappa_{\tilde{t}_i}\right)} T^{t_{-i}[a_{-i}]}_i \right), \\
T^{\tilde{t}_i}_j &= \bigcup_{(\theta, t_{-i}[a_{-i}]) \in\, \mathrm{supp}\left(\kappa_{\tilde{t}_i}\right)} T^{t_{-i}[a_{-i}]}_j \qquad (j \neq i).
\end{aligned}$$

It is straightforward to show that $\mathrm{supp}\big(\kappa_{\tilde{t}_i}\big)$ finite, and hence $T^{\tilde{t}_i}$ is finite and belief-closed (see Yildiz (2005)). Finally, since $S^{k+1}_i\left[\tilde{t}_i\right] = \{a_i\}$, $\left|S^{k+1}_i\left[\tilde{t}_i\right]\right| = 1$, and by construction, for each $t_j \in T^{\tilde{t}_i}_j \setminus \left\{\tilde{t}_i\right\}$, $\left|S^{k+1}\left[t_j\right]\right| = \left|S^{k}\left[t_j\right]\right| = 1$.

To prove the last statement in the lemma, take any $a_i \in W^{\infty}_i\left[\hat{t}_i\right]$. For each m, since $a_i \in W^{\infty}_i\left[\hat{t}_i\right] \subseteq W^{m}_i\left[\hat{t}_i\right]$, by the first part of the lemma, there exists $t_{i,m}$ such that $t^k_{i,m} = \hat{t}^k_i$ for each $k \leq m$ and $S^{m+1}_i\left[t_{i,m}\right] = S^{\infty}_i\left[t_{i,m}\right] = \{a_i\}$. Clearly, for any fixed k, $t^k_{i,m} = \hat{t}^k_i$ for each $m > k$, showing that $t^k_{i,m} \to \hat{t}^k_i$ as $m \to \infty$. By the first part, $t_{i,m} \in T^{t_{i,m}}_i$ for some finite model $T^{t_{i,m}}$ with $\left|S^{\infty}\left[t\right]\right| = \left|S^{m+1}\left[t\right]\right| = 1$ for each $t \in T^{t_{i,m}}$. Pick $T^m = T^{t_{i,m}}$ as the dominance-solvable model in the lemma. □

Finally, we can combine Lemmas 6 and 7 to conclude that we can perturb the beliefs slightly in the universal type space to make any rationalizable action strictly rationalizable:

Proof of Proposition 1. Take any $\hat{t} \in \hat{T}$, and any $a \in S^{\infty}\left[\hat{t}\right]$. By Lemma 6, there exists a sequence $\bar{t}(m) \in \hat{T}$ such that $a \in W^{\infty}\left[\bar{t}(m)\right]$ and $\bar{t}(m) \to \hat{t}$ as $m \to \infty$. But by Lemma 7, since $a \in W^{\infty}\left[\bar{t}(m)\right]$, for each m and k, there exists a finite, dominance-solvable model $T^{m,k}$ with a type profile $t(m,k)$, such that $S^{\infty}\left[t(m,k)\right] = \{a\}$ and $t(m,k) \to \bar{t}(m)$ as $k \to \infty$. Since T^* is metrizable, there then exists a sequence $k_m \to \infty$ with $t(m,k_m) \to \hat{t}$. Set $\tilde{t}(m) = t(m,k_m)$ and $T^m = T^{m,k_m}$, which satisfy the desired properties. □

A.3. *Proof of Proposition* 4

We will construct a finite model $\tilde{T}^{s_T,m}$ that admits a common prior with full support and has the desired properties. Since T^{CPA} is dense, for each $\tau(t,s_T,m)$ in Proposition 3, there exists a sequence of finite models $T^{t,m,k} \subset \hat{T}$ with common priors $p^{t,m,k}$ (with full support) and members $\bar{\tau}(t,m,k)$ such that $\bar{\tau}(t,m,k) \to \tau(t,s_T,m)$ as $k \to \infty$. By the last statement in Proposition 2, there exists a $\bar{k}$ such that for each $k > \bar{k}$, $S^{\infty}\left[\bar{\tau}(t,m,k)\right] = S^{\infty}\left[\tau(t,m,k)\right] = \{s_T(t)\}$. Hence, without loss of generality, pick each $\bar{\tau}(t,m,k)$ with

$$S^{\infty}\left[\bar{\tau}(t,m,k)\right] = \{s_T(t)\}. \tag{A.5}$$

For each $\varepsilon \in [0,1]$, we will now construct a finite model $\left(\Theta \times T^{m,k,\varepsilon}, p^{m,k,\varepsilon}\right)$ with common prior $p^{m,k,\varepsilon}$ in which the types are denoted by integers. For each i, let $\hat{\tau}_i$ be any one-to-one mapping that maps types $\tilde{t}_i$, $\tilde{t}_i \in T_i^{t,m,k}$, $t \in T$, to integers. (Recall that there are only finitely many such types.) Define

$$T_i^{m,k,\varepsilon} = \left\{\hat{\tau}_i\left(\tilde{t}_i\right) | \tilde{t}_i \in T_i^{t,m,k}, t \in T\right\} \qquad (i \in N).$$

Let Θ be the set of all $\theta \in \Theta^*$ on which some type $\tilde{t}_i \in T_i^{t,m,k}$ puts positive probability. We will now define a common prior $p^{m,k,\varepsilon}$ on $\Theta \times T^{m,k,\varepsilon}$ with full support. Since each $p^{t,m,k}$ has full support, given any t and t', either $T^{t,m,k} = T^{t',m,k}$ or $T^{t,m,k} \cap T^{t',m,k} = \varnothing$. Let K be the number of disjoint sets $T^{t,m,k}$ and $L = \left|\Theta \times T^{m,k,\varepsilon}\right|$. Define $p^{m,k,\varepsilon}$ by setting

$$p^{m,k,\varepsilon}\left(\theta,\bar{t}\right) = \begin{cases} \varepsilon/L + (1-\varepsilon)\, p^{t,m,k}\left(\theta,\tilde{t}\right)/K & \text{if } \bar{t} = \hat{\tau}\left(\tilde{t}\right) \text{ for some } \tilde{t} \in T^{t,m,k} \text{ and } t, \\ \varepsilon/L & \text{otherwise} \end{cases}$$

at each $\left(\theta,\bar{t}\right) \in \Theta \times T^{m,k,\varepsilon}$. According to $p^{m,k,\varepsilon}$, with probability ε, we have a uniform distribution on $\Theta \times T^{m,k,\varepsilon}$, and with probability $(1-\varepsilon)$ one of the type spaces $T^{t,m,k}$ is selected, each with equal probability. Let $h_i\left(\hat{\tau}_i\left(\tilde{t}_i\right); m,k,\varepsilon\right)$ be the belief hierarchy of $\hat{\tau}_i\left(\tilde{t}_i\right)$ under $p^{m,k,\varepsilon}$. Applying Lemma 4 to the model that consists of submodels $\left(\Theta \times T^{m,k,\varepsilon}, p^{m,k,\varepsilon}\right)$, $\varepsilon \in [0,1]$, we conclude that as $\varepsilon \to 0$, $h_i\left(\hat{\tau}_i\left(\tilde{t}_i\right); m,k,\varepsilon\right) \to h_i\left(\hat{\tau}_i\left(\tilde{t}_i\right); m,k,0\right)$ for each $\hat{\tau}_i\left(\tilde{t}_i\right)$. Moreover, $h_i\left(\hat{\tau}_i\left(\tilde{t}_i\right); m,k,0\right) = \tilde{t}_i$ by construction. (A type's belief hierarchy cannot change when we add an ex ante stage to choose between type spaces.) Therefore, $h_i\left(\hat{\tau}_i\left(\tilde{t}_i\right); m,k,\varepsilon\right) \to \tilde{t}_i$ for each $\tilde{t}_i$. This implies by the last statement in Proposition 2 and (A.5) that, for each

(t, m, k), there exists $\varepsilon^{t,m,k} > 0$ such that $S^{\infty}\left[h\left(\hat{\tau}\left(\bar{\tau}\left(t, m, k\right)\right), m, k, \varepsilon\right)\right] = \left\{s_T\left(t\right)\right\}$ whenever $\varepsilon < \varepsilon^{t,m,k}$. Since T^* is metrizable, there then exist sequences $k_m \to \infty$ and $\varepsilon_m \to 0$ with $\varepsilon_m < \min_{t \in T} \varepsilon^{t,m,k_m}$ such that $h\left(\hat{\tau}\left(\bar{\tau}\left(t, m, k_m\right)\right), m, k_m, \varepsilon_m\right) \to t$ as $m \to \infty$. For each t and m, set $\tilde{T}^{s_T,m} = h\left(T^{m,k_m,\varepsilon_m}; m, k_m, \varepsilon_m\right)$ and $\tilde{\tau}\left(t, s_T, m\right) = h\left(\hat{\tau}\left(\bar{\tau}\left(t, m, k_m\right)\right), m, k_m, \varepsilon_m\right)$. Since $\tilde{\tau}\left(t, s_T, m\right) \to t$ for each t, we can pick ε_m such that $\tilde{\tau}\left(t, s_T, m\right)$ is one-to-one. By construction, $S^{\infty}\left[\tilde{\tau}\left(t, s_T, m\right)\right] = \left\{s_T\left(t\right)\right\}$ for each (t, m), and $\tilde{\tau}\left(t, s_T, m\right) \to t$ as $m \to \infty$.

A.4. *Generic Uniqueness in Rich Models*

In this paper, we use the product topology on the universal type space. Using Corollary 1 and Lemma 4, we will next show that our genericity result holds for any type space with respect to its own topology, provided that it satisfies the technical conditions in Lemma 4 and it is rich enough so that it generates a dense set of belief hierarchies. Indeed, it turns out that any such model is isomorphic to the universal type space with product topology, which emphasizes the central nature of this space and topology.

Corollary 3. *Let $M = (\Theta \times T, \kappa)$ be any model, endowed with any topology, such that (i) $\Theta \times T$ is compact, (ii) κ_{t_i} is a continuous function of t_i, (iii) $(\Theta \times T, \kappa)$ does not have redundant types, and (iv) $h\left(T\right)$ is dense. Then, under Assumption 1,*

$$U^M = \{t \in T | t \textit{ has a unique rationalizable action profile}\}$$

is open and dense with respect to the topology on T.

Proof. By definition, $h\left(U^M\right) = U \cap h\left(T\right)$. Since $h\left(T\right)$ is dense, Corollary 1 implies that $h\left(U^M\right)$ is dense and open with respect to the relative topology on $h\left(T\right)$. By Lemma 4, $h : T \to h\left(T\right)$ is continuous. Since $h\left(U^M\right)$ is open relative to $h\left(T\right)$, this implies that $U^M = h^{-1}\left(h\left(U^M\right)\right)$ is open. By Lemma 4, h^{-1} is also continuous and onto. Since $h\left(U^M\right)$ is dense, this implies that $U^M = h^{-1}\left(h\left(U^M\right)\right)$ is also dense. □

REFERENCES

Aumann, R. (1987): Correlated Equilibrium as an Expression of Bayesian Rationality, *Econometrica*, 55, 1-18.

Battigalli, P. (2003): Rationalizability in Infinite, Dynamic Games with Complete Information, *Research in Economics*, 57, 1-38.

Battigalli, P. and M. Siniscalchi (2003) Rationalization and Incomplete Information, *Advances in Theoretical Economics*, 3-1, Article 3, http://www.bepress.com.

Bernheim, D. (1984): Rationalizable Strategic Behavior, *Econometrica*, 52, 1007-1028.

Borgers T. (1994): Weak Dominance and Approximate Common Knowledge, *Journal of Economic Theory*, 64, 265-276.

Brandenburger, A. and E. Dekel (1987): Rationalizability and Correlated Equilibria, *Econometrica*, 55, 1391-1402.

——— (1993): Hierarchies of Beliefs and Common Knowledge, *Journal of Economic Theory*, 59, 189-198.

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

Cremer, J. and. R. McLean (1988): Full extraction of the Surplus in Bayesian and Dominant Strategy Auctions, *Econometrica*, 56, 1247-1257.

Dekel, E. and D. Fudenberg (1990): Rational Behavior with Payoff Uncertainty, *Journal of Economic Theory*, 52, 243-267.

Dekel, E., D. Fudenberg, S. Morris (2003): Interim Rationalizability, Unpublished Manuscript, Harvard University.

——— (2006): Topologies on Types, *Theoretical Economics* 1, 275–309, http://econtheory.org.

Ely, J. and M. Peski (2006): Hierarchies of Beliefs and Interim Rationalizability, *Theoretical Economics* 1, 275–309, http://econtheory.org.

Feinberg, Y. (2000): Characterizing Common Priors in the Form of Posteriors, *Journal of Economic Theory*, 91, 127-179.

Feinberg, Y. and A. Skrzypacz (2005): Uncertainty about Uncertainty and Delay in Bargaining, *Econometrica*, 73, 69-91.

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

Fudenberg, D., D. Kreps, and D. Levine (1988): On the Robustness of Equilibrium Refinements, *Journal of Economic Theory* 44, 354-380.

Harsanyi, J. (1967): Games with Incomplete Information Played by Bayesian Players. Part I: The Basic Model, *Management Science* 14, 159-182.

Harsanyi, J. and R. Selten (1988): *A General Theory of Equilibrium Selection in Games.* Cambridge, MA: The MIT Press.

Heifetz, A. and Z. Neeman (2006): On the Generic (Im)Possibility of Full Surplus Extraction in Mechanism Design, *Econometrica*, 74, 213-234.

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

Kreps, D., P. Milgrom, J. Roberts and R. Wilson (1982): Rational Cooperation in the Finitely-Repeated Prisoners' Dilemma, *Journal of Economic Theory*, 27, 245-52.

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

Mertens, J. and S. Zamir (1985): Formulation of Bayesian Analysis for Games with Incomplete Information, *International Journal of Game Theory*, 10, 619-632.

Monderer, D. and D. Samet (1989): Approximating Common Knowledge with Common Beliefs, *Games and Economic Behavior*, 1, 170-190.

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

Morris, S. and H. Shin (2000): Rethinking Multiple Equilibria in Macroeconomics, *NBER Macroeconomics Annual*, 139-161.

Neeman, Z. (2004) The Relevance of Private Information in Mechanism Design, *Journal of Economic Theory*, 117, 55-77.

Oxtoby, J. (1980): *Measure and Category*, Berlin: Springer-Verlang.

Oyama, D. and O. Tercieux (2005): Strategic Implications of (Non-)Common Priors II: Robustness of Equilibria, Unpublished Manuscript, University of Tokyo.

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

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

Simon, R. (2003): Games of Incomplete Information, Ergodic Theory, and the Measurability of Equilibria, *Israeli Journal of Mathematics* 138, 73-92.

Weinstein, J. and M. Yildiz (2004): Finite-order Implications of Any Equilibrium, *MIT Department of Economics Working Paper* 03-14.

Wilson, R. (1987): Game-Theoretic Analyses of Trading Processes, in: Truman Bewley (ed.) *Advances in Economic Theory: Fifth World Congress*, Cambridge UK: Cambridge University Press, 33-70.

Yildiz, M. (2005): Generic Uniqueness of Rationalizable Actions, Unpublished Manuscript, MIT.