**INFORMATIONAL ROBUSTNESS AND SOLUTION CONCEPTS**


**By**

**Dirk Bergemann and Stephen Morris**



**December 2014**

**COWLES FOUNDATION DISCUSSION PAPER NO. 1973**



**COWLES FOUNDATION FOR RESEARCH IN ECONOMICS**
**YALE UNIVERSITY**
**Box 208281**
**New Haven, Connecticut 06520-8281**

**http://cowles.econ.yale.edu/**

# Informational Robustness and Solution Concepts*

Dirk Bergemann[†] Stephen Morris[‡]

First Version: January 2014
Current Version: December 15, 2014

**Abstract**

We discuss four solution concepts for games with incomplete information. We show how each solution concept can be viewed as encoding informational robustness. For a given type space, we consider *expansions* of the type space that provide players with additional signals. We distinguish between expansions along two dimensions. First, the signals can either convey payoff relevant information or only payoff irrelevant information. Second, the signals can be generated from a common (prior) distribution or not. We establish the equivalence between Bayes Nash equilibrium behavior under the resulting expansion of the type space and a corresponding more permissive solution concept under the original type space. This approach unifies some existing literature and, in the case of an expansion without a common prior and allowing for payoff relevant signals, leads us to a new solution concept that we dub belief-free rationalizability.

Jel Classification: C79, D82.

Keywords: Incomplete Information, Informational Robustness, Bayes Correlated Equilibrium, Interim Corrrelated Rationalizability, Belief Free Rationalizability

*We acknowledge financial support from NSF ICES 1215808. We are grateful for discussions about an earlier version of this paper with Pierpaolo Battigalli, Eddie Dekel, Qingmin Liu and Marciano Siniscalchi.

[†]Department of Economics, Yale University, New Haven, U.S.A., dirk.bergemann@yale.edu.

[‡]Department of Economics, Princeton University, Princeton, U.S.A., smorris@princeton.edu.

# 1 Introduction

Classical analysis of incomplete information games treats the information structure of the players as given, and examines the consequences of some theory of rational behavior (i.e., solution concept) given that information structure. But the exact information structure is often not known to the analyst, and thus it is interesting to examine the implications of that solution concept in all information structures that the analyst thinks possible, and thus identify predictions that are robust to informational assumptions. In earlier work, we have examined such informational robustness questions both in the context of mechanism design (Bergemann and Morris (2012)) and in the context of general games (Bergemann and Morris (2014)).

There is a close connection between relaxing informational assumptions and relaxing solution concepts. Consider the solution concept of Nash equilibrium in a complete information game. Suppose that we allow players to observe arbitrary (payoff-irrelevant) signals. If the common prior assumption is maintained, then Aumann (1987) showed that distribution of equilibrium behavior would correspond to an (objective) correlated equilibrium. Without the common prior assumption, Brandenburger and Dekel (1987) and Tan and Werlang (1988) showed that all one can say about the resulting equilibrium behavior is that each player will choose a (correlated) rationalizable action.

What are the incomplete information analogues of these results? Suppose now that payoffs depend on a "payoff state" and players' beliefs and higher-order beliefs about that state are described by a type space. But suppose that players may also observe payoff-irrelevant signals that do not change their beliefs and higher-order beliefs about the state. Incomplete information analogues of the complete information results are known in this setting. If the common prior assumption is maintained, and we study (Bayes Nash) equilibria on the expanded type space with payoff-irrelevant signals, then the distribution of equilibrium behavior corresponds to a *belief invariant Bayes correlated equilibrium* (Liu (2014)). Without the common prior assumption, all one can say about the resulting equilibrium behavior is that players will choose *interim correlated rationalizable* actions (Dekel, Fudenberg, and Morris (2006)). Alternative extensions of the complete information results to incomplete information arise if we allow players to observe payoff *relevant* signals, i.e., signals that refine their initial beliefs and higher-order beliefs about the state. If the common prior assumption is maintained, and we study equilibria on the expanded type space with payoff-relevant signals, then the distribution of equilibrium behavior corresponds to a *Bayes correlated equilibrium* (Bergemann and Morris (2014)).

This summary of existing incomplete information results leaves one open question: what is the implication of allowing players to observe additional payoff relevant signals without imposing the common prior assumption? One contribution of this paper is describe a solution concept - *belief-free rationalizability* - and show that an action can be played in equilibrium by a given type who may observe extra payoff relevant signals if and only if it is belief-free rationalizable.[1] Belief-free rationalizability is defined by the following iterative process that uses only the *support* of a type's beliefs, i.e., the set of profiles of other players' types and states that he thinks possible. At each round, we delete an action for a particular type if there is no conjecture over profiles of other players' actions and states, such that the action is a best response to the conjecture; and the conjecture assigns zero probability to (1) profiles of the other players' actions and types that have already been deleted; and (2) profiles of other players' types and states that are not in the support of his original beliefs.

The following table summarizes the consequences of equilibrium under incomplete information if we allow players to observe additional signals that may or may not be consistent with the common prior and may or may not be payoff relevant:

| | payoff relevant signals | payoff-irrelevant signals only |
|---|---|---|
| common prior | Bayes correlated equilibrium | belief invariant Bayes correlated equilibrium |
| non common prior | belief-free rationalizability | interim correlated rationalizability |

In the special case of complete information (i.e., a unique payoff state), both rationalizability results reduce to the result of Brandenburger and Dekel (1987) and both correlated equilibrium results reduce to the result of Aumann (1987).

We report one example to illustrate belief-free rationalizability. Consider a two player two action game where each player must decide whether to invest or not. There is a bad state where it is a dominant strategy for both players to not invest. But there is also a good state where invest is a best response for a player only if he assigns at least probability $p$ to the other player investing. Investment is interim correlated rationalizable in this example only if there is common $p$-belief that the state is good. This also implies that the largest belief invariant Bayes correlated equilibrium has both players investing only when there is common $p$-belief that the state is good. We give a parallel characterization for belief-free rationalizability. A necessary

[1] We discuss the relation to the use of this name in Battigalli, Di Tillio, Grillo, and Penta (2011) below.

condition for investment will be that a player thinks that it is possible that the state is good. Another necessary condition is that a player thinks it is possible that the other player thinks that it is possible that the state is good. And so on. Thus a necessary - and, one can also show, sufficient - condition for investment to be belief-free rationalizable is that there is "common possibility" that the state is good, in the sense that all infinite sequences of such statements are true. For an insightful equivalent characterization, say that there is iterated distributed certainty of an event if someone is sure that the event is true, or someone is sure that someone else is sure that the event is true, or some iteration of this statement is true. One can also show that invest is belief-free rationalizable if and only if there is not distributed iterated certainty that the state is bad.

All these solution concepts have simpler statements and interpretations in the special case of "payoff type" environments, where we assume the payoff state can be represented as a profile of player specific "payoff types", and each player is certain of his own payoff type. The payoff type assumption corresponds to the assumption that there is "distributed certainty," i.e., the join of players' information reveals the true state; the assumption is not without loss of generality. But under this assumption, the "correlation" in interim correlation rationalizability is no longer relevant, and it is equivalent to interim independent rationalizability; the belief invariant Bayes correlated equilibrium reduces to the belief invariant Bayesian solution of Forges (2006) and Lehrer, Rosenberg, and Shmaya (2010); and Bayes correlated equilibrium reduces to the Bayesian solution of Forges (1993).

Much of the literature - for one reason or another - focusses on the special case of "payoff type" environments. This assumption is implicit in much of the literature on incomplete information correlated equilibrium, e.g., in the solution concepts and papers cited in the previous paragraph. A leading example of a payoff type environment is a private values environment (where a player's payoff depends only on his own payoff type), and Chen, Micali, and Pass (2014b) have proposed what we are calling belief-free rationalizability in this context and used it for novel results on robust revenue maximization in Chen, Micali, and Pass (2014a). Payoff type environments without private values were the focus of earlier work of ours on robust mechanism design collected in Bergemann and Morris (2012). In that work, we considered the special case where all players believed that all payoff type profiles of other players were possible. In this special case, belief-free rationalizability has a particularly simple characterization. For each payoff type, iteratively delete actions for that payoff type which are not a best response against any conjecture over others players' actions and types that have survived the iterated deletion procedure so far. This solution concept (with appropriate informational robustness foundations) was used in a number of our papers on robust mechanism design (i.e., chapters 3, 4 and 7 of Bergemann and Morris (2012)). Battigalli, Di Tillio, Grillo, and Penta (2011) labelled this solution concept "belief-free rationalizability". We used Bayes correlated equilibrium (in the special case of payoff type environments) in Bergemann and Morris (2008) (chapter 5 in Bergemann and Morris (2012)).[2] A second contribution of this paper is then to tightly relate our earlier work on robust mechanism design to our more recent work on robust predictions in games (Bergemann and Morris (2013), (2014)).

Battigalli and Siniscalchi (2003b) introduced the notion of "$\Delta$-rationalizability" for both complete and incomplete information environments, building in arbitrary restrictions on the beliefs of any type about other players' types and actions, and states. Battigalli, Di Tillio, Grillo, and Penta (2011) describes how interim correlated rationalizability (in general) and belief-free rationalizability (in the case of payoff type spaces) are special cases of "$\Delta$-rationalizability", where particular restrictions are placed on beliefs about other players' types and states. Belief-free rationalizability will also be a special case of $\Delta$-rationalizability, outside payoff type environments, where the corresponding type dependent restriction on beliefs would be on the support of beliefs only.

We framed the complete information results of Brandenburger and Dekel (1987) and Aumann (1987) as "informational robustness" results, i.e., what happens to equilibrium predictions if we allow players to observe additional (payoff-irrelevant) signals, and this corresponds to the formal statements of their results.[3] However, both papers interpret their results informally as establishing the implications of common certainty of rationality,[4] and the later "epistemic foundations"

[2] Thus the application of the solution concepts in the results in this paper to payoff type environments subsumes results we reported in our unpublished paper on "Belief Free Incomplete Information Games" (Bergemann and Morris (2007)).

[3] Thus Proposition 2.1 of Brandenburger and Dekel (1987), while stated in the language of interim payoffs, established that the set of actions played in an appropriate version of subjective correlated equilibrium were equal to the correlated rationalizable actions. The main theorem of Aumann (1987) showed that under assumptions equivalent to Bayes Nash equilibrium on a common prior type space with payoff-irrelevant signals, the ex ante distribution of play corresponds to an (objective) correlated equilibrium.

[4] Aumann (1987) notes in the introduction that he assumes "common knowledge that each player chooses a

literature presents more formal statements of these results as consequence of common certainty of rationality.[5] We also discuss how our "informational robustness" results can be translated back into "epistemic foundations" results, justifying solution concepts from primitive epistemic assumptions. However, our epistemic foundations results are in the spirit of the classical literature and we do not address issues that have been the focus of much recent literature, i.e., removing reference to players' beliefs about their own types (Aumann and Brandenburger (1995)), restricting attention to state spaces that reflect "expressible" statements about the model (Brandenburger and Friedenberg (2008) and Battigalli, Di Tillio, Grillo, and Penta (2011)), and giving an interim interpretation of the common prior results (Dekel and Siniscalchi (2014)). The focus of our discussion of epistemic foundations is showing how the informational robustness results we discuss map into the modern epistemic foundations literature, without attempting a treatment of the topics of interest in that literature.

We discuss the four solution concepts in section 2 and the coordination example in section 3. In section 4, we report how the solution concepts specialize to complete information rationalizability and correlated equilibrium in the case of complete information games, and widely used and simpler solution concepts in the case of payoff type environments. Informational robustness foundations of the solution concepts are reported in section 5; their translation to (old fashioned) epistemic foundation results and their relation to the (modern) epistemic foundations literature are discussed in section 6.

---

strategy that maximizes his expected utility given his information". Brandenburger and Dekel (1987) write in the introduction that their approach "starts from the assumption that the rationality of the players is common knowledge." We follow the recent literature in replacing the term "knowledge" in the expression common knowledge because it corresponds to "belief with probability 1," rather than "true belief" (the meaning of knowledge in philosophy and general discourse). We use "certainty" to mean "belief with probability 1".

[5] Thus Dekel and Siniscalchi (2014) state a modern version of the main result of Brandenburger and Dekel (1987) as Theorem 1 and a (somewhat) more modern statement of Aumann (1987) in section 4.6.2.

# 2 Four Classical Solution Concepts

We will fix a finite set of players $1, ..., I$ and a finite set of payoff relevant states $\Theta$.

We divide a standard description of an incomplete information game into a "basic game" and a "type space". A basic game $\mathcal{G} = (A_i, u_i)_{i=1}^{I}$ consists of, for each player, a finite set of possible actions $A_i$ and a payoff function $u_i : A \times \Theta \to \mathbb{R}$ where $A = A_1 \times ... \times A_I$.[6] A type space $\mathcal{T} = (T_i, \pi_i)_{i=1}^{I}$ consists of, for each player, a finite set of types, $T_i$ and, for each of his types, a belief over the others players' types and the state, $\pi_i : T_i \to \Delta(T_{-i} \times \Theta)$. An incomplete information game consists of a basic game $\mathcal{G} = (A_i, u_i)_{i=1}^{I}$ and a type space $\mathcal{T} = (T_i, \pi_i)_{i=1}^{I}$; this is the standard description modulo the fact that the common prior assumption is not maintained. We begin by discussing "classical solution concepts" for the fixed incomplete information game $(\mathcal{G}, \mathcal{T})$, meaning that we define solution concepts without referring to informational robustness (or epistemic) foundations.

We consider two alternative definitions of rationalizability in game $(\mathcal{G}, \mathcal{T})$. First consider *interim correlated rationalizability* (Dekel, Fudenberg, and Morris (2007)). An action is interim correlated rationalizable for a type $t_i$ if we iteratively delete actions which are not a best response to any supporting conjecture over other players' actions and types, as well as states, which (1) puts probability 1 on action type profiles which have survived the iterated deletion procedure so far, and (2) has a marginal belief over others' types and states which is consistent with that type's beliefs on the type space. Crucially, this definition allows arbitrary correlation in the supporting conjecture as long as (1) and (2) are satisfied. Formally, let $ICR_i^0(t_i) = A_i$ and let $ICR_i^{n+1}(t_i)$ equal the set of actions for which there exists $\nu_i \in \Delta(A_{-i} \times T_{-i} \times \Theta)$ such that

$$(1)\ \nu_i(a_{-i}, t_{-i}, \theta) > 0 \Rightarrow a_j \in ICR_j^n(t_j) \text{ for each } j \neq i;$$
$$(2)\ \sum_{a_{-i}} \nu_i(a_{-i}, t_{-i}, \theta) = \pi_i(t_{-i}, \theta | t_i) \text{ for each } t_{-i}, \theta;$$
$$(3)\ a_i \in \arg\max_{a_i'} \sum_{a_{-i}, t_{-i}, \theta} \nu_i(a_{-i}, t_{-i}, \theta)\, u_i((a_i', a_{-i}), \theta);$$

and let

$$ICR_i(t_i) = \bigcap_{n \geq 1} ICR_i^n(t_i).$$

[6] In Bergemann and Morris (2014) we included a common prior on states in the description of the basic game. Because we are relaxing the common prior assumption, it is convenient to use a slightly different definition in this paper.

**Definition 1 (Interim Correlated Rationalizable)**
*Action $a_i$ is interim correlated rationalizable for type $t_i$ (in game $(\mathcal{G}, \mathcal{T})$) if $a_i \in ICR_i(t_i)$.*

Now consider a more permissive rationalizability notion, *belief-free rationalizability.* The definition is the same as iterated correlated rationalizability except that we relax assumption (2) to the requirement that the rationalizing conjecture be consistent with the player's belief on the type space to the weaker requirement that its support is a subset of the player's belief on the type space. Thus we have $BFR_i^0(t_i) = A_i$ and let $BFR_i^{n+1}(t_i)$ equal the set of actions for which there exists $\nu_i \in \Delta(T_{-i} \times A_{-i} \times \Theta)$ s.t.

$$\begin{aligned}
&(1)\ \nu_i(a_{-i}, t_{-i}, \theta) > 0 \Rightarrow a_j \in BFR_j^n(t_j) \text{ for each } j \neq i; \\
&(2)\ \sum_{a_{-i}} \nu_i(a_{-i}, t_{-i}, \theta) > 0 \Rightarrow \pi_i(t_{-i}, \theta | t_i) > 0 \text{ for each } t_{-i}, \theta; \\
&(3)\ a_i \in \arg\max_{a_i'} \sum_{a_{-i}, t_{-i}, \theta} \nu_i(a_{-i}, t_{-i}, \theta)\, u_i((a_i', a_{-i}), \theta);
\end{aligned}$$

and let

$$BFR_i(t_i) = \bigcap_{n \geq 1} BFR_i^n(t_i).$$

**Definition 2 (Belief-Free Rationalizable)**
*Action $a_i$ is belief-free rationalizable for type $t_i$ (in game $(\mathcal{G}, \mathcal{T})$) if $a_i \in BFR_i(t_i)$.*

Note that this definition is independent of a type's quantized beliefs and depends only on which profiles of other players' types and states he considers possible, i.e., the support of his beliefs.

We now consider two parallel definitions of (objective) incomplete information correlated equilibrium for the same incomplete information game. Type space $\mathcal{T} = (T_i, \pi_i)_{i=1}^I$ satisfies the *common prior assumption* if there exists $\pi^* \in \Delta(T \times \Theta)$ such that

$$\sum_{t_{-i}', \theta'} \pi^*\left(\left(t_i, t_{-i}'\right), \theta'\right) > 0$$

for all $i$ and $t_i$, and

$$\pi_i(t_{-i}, \theta | t_i) = \frac{\pi^*((t_i, t_{-i}), \theta)}{\sum\limits_{t_{-i}', \theta'} \pi^*\left(\left(t_i, t_{-i}'\right), \theta'\right)}$$

for all $i$, $(t_i, t_{-i})$ and $\theta$.[7]

Now we have a common prior incomplete information game $(\mathcal{G}, \mathcal{T})$. Behavior in this incomplete information game can be described by a *decision rule* mapping players' types and states to a probability distribution over players' actions, $\sigma : T \times \Theta \to \Delta(A)$. A decision rule $\sigma$ satisfies *belief invariance* if, for each player $i$,

$$\sigma_i\left(a_i | \left(t_i, t_{-i}\right), \theta\right) \triangleq \sum_{a_{-i}} \sigma_i\left(\left(a_i, a_{-i}\right) | \left(t_i, t_{-i}\right), \theta\right)$$

is independent of $(t_{-i}, \theta)$. Thus a decision rule satisfies belief invariance if a player's action recommendation does not reveal any additional information to him about others' types and the state. This property has played an important role in the literature on incomplete information correlated equilibrium, see, Forges (1993), Forges (2006) and Lehrer, Rosenberg, and Shmaya (2010). Notice that property (2) in the iterative definition of interim correlated rationalizability was a belief invariance assumption.

Decision rule $\sigma$ satisfies *obedience* if

$$\begin{aligned} & \sum_{a_{-i}, t_{-i}, \theta} \pi^*\left(t_i, t_{-i}\right) \sigma\left(\left(a_i, a_{-i}\right) | \left(t_i, t_{-i}\right), \theta\right) u_i\left(\left(a_i, a_{-i}\right), \theta\right) \\ \geq & \sum_{a_{-i}, t_{-i}, \theta} \pi^*\left(t_i, t_{-i}\right) \sigma\left(\left(a_i, a_{-i}\right) | \left(t_i, t_{-i}\right), \theta\right) u_i\left(\left(a_i', a_{-i}\right), \theta\right). \end{aligned}$$

for all $i$, $t_i \in T_i$ and $a_i, a_i' \in A_i$. Obedience has the following mediator interpretation. Suppose that an omniscient mediator knew players' types and the true state, randomly selected an action profile according to $\sigma$ and privately informed each player of his recommended action. Would a player who knew his own type and heard the mediator's recommendation have an incentive to follow the recommendation? Obedience says that he would want to follow the recommendation.

**Definition 3 (Belief Invariant Bayes Correlated Equilibrium)**
*Decision rule $\sigma$ is a belief invariant Bayes correlated equilibrium if it satisfies obedience and belief invariance.*

[7] When the common prior assumption is maintained, we understand the common prior $\pi^*$ to be implicitly defined by the type space. In the (special) case where multiple common priors satisfy the above properties, our results will hold true for any choice of common prior. By requiring that all types are assigned positive probability, we are making a slightly stronger assumption than some formulations of results in the literature. This version simplifies the statement of results and will also tie in with the support assumption that we impose in the informational robustness foundations in Section 5.

Liu (2014) described the subjective correlated equilibrium analogue of interim correlated rationalizability. If one then imposes the common prior assumption (as he discusses in section 5.3), this is the version of incomplete information correlated equilibrium one obtains. Its relation to the incomplete information correlated equilibrium literature is further discussed in Bergemann and Morris (2014): it is in general a weaker requirement than the belief invariant Bayesian solution of Forges (2006) and Lehrer, Rosenberg, and Shmaya (2010), because - like interim correlated rationalizability - it allows unexplained correlation between types and payoff states. It is immediate from definitions that any action played with positive probability by a type in a belief invariant Bayes correlated equilibrium is interim correlated rationalizable.

**Definition 4 (Bayes Correlated Equilibrium)**
*Decision rule $\sigma$ is a Bayes correlated equilibrium if it satisfies obedience.*

This solution concept is studied in Bergemann and Morris (2014). It is immediate from definitions that any action played with positive probability by a type in a belief invariant Bayes correlated equilibrium is belief-free rationalizable.

# 3 Binary Action Coordination Games

The solution concepts can be illustrated by a classic binary action coordination game. There are two states, "good" (G) and "bad" (B), so $\Theta = \{\theta_G, \theta_B\}$. There are two actions, Invest and Not Invest. Payoffs are given by the following matrices:

| $\theta_B$ | Invest | Not Invest |
|---|---|---|
| Invest | $-1, -1$ | $-1, 0$ |
| Not Invest | $0, -1$ | $0, 0$ |

| $\theta_G$ | Invest | Not Invest |
|---|---|---|
| Invest | $\frac{1}{p} - 1, \frac{1}{p} - 1$ | $-1, 0$ |
| Not Invest | $0, -1$ | $0, 0$ |

where $0 < p < 1$. Thus there is a dominant strategy to not invest in the bad state. There are two strict Nash equilibria in the good state. If there is common certainty that the state is good, invest is a best response for a player if and only if he thinks the other player will invest with probability at least $p$.

We will characterize rationalizable behavior (for the two versions given above) on all type spaces using belief operators. For the fixed type space, an event $E$ is a subset of $T \times \Theta$. Following

Monderer and Samet (1989), we define belief operators as follows. Write $B_i^p(E)$ for set of types of player $i$ who believe event $E$ with probability at least $p$, so

$$B_i^p(E) = \left\{ t_i \in T_i \left| \pi_i \left( \left\{ \left(t'_{-i}, \theta'\right) \mid \left( \left(t_i, t'_{-i}\right), \theta' \right) \in E \right\} | t_i \right) \geq p \right. \right\}.$$

Also define everyone $p$-believes and common $p$-belief operators which map events in $T \times \Theta$ to other events in $T \times \Theta$:

$$B_*^p(E) = \left( \underset{i=1,\dots,I}{\times} B_i^p(E) \right) \times \Theta \text{ and } C^p(E) = \bigcap_{n=1,2,\dots} \left[B_*^p\right]^n(E).$$

Write $E_G$ and $E_B$ for the set of states where the payoff state $\theta$ is good and bad respectively, so

$$E_G = \{(t,\theta) | \theta = \theta_G\} \text{ and } E_B = \{(t,\theta) | \theta = \theta_B\}.$$

Now action "not invest" is always interim correlated rationalizable. Action "invest" is interim correlated rationalizable for type $t_i$ of player $i$ only if he $p$-believes that it is common $p$-belief that the state is good, i.e., if $t_i \in B_i^p(C^p(E_G))$. This is a well known characterization.[8]

Now write $B_i^+(E)$ for the set of types of player $i$ that think that $E$ is possible:

$$B_i^+(E) = \left\{ t_i \in T_i \left| \begin{array}{l} \exists t_{-i} \in T_{-i} \text{ and } \theta \in \Theta \text{ such that} \\ \left( \left(t_i, t_{-i}\right), \theta \right) \in E \text{ and } \pi_i\left(t_{-i}, \theta | t_i\right) > 0 \end{array} \right. \right\} = \underset{p>0}{\cap} B_i^p(E)$$

We can define everyone thinks possible and common possibility operators in the natural ways:

$$\begin{aligned} B_*^+(E) &= \left( \underset{i=1,\dots,I}{\times} B_i^+(E) \right) \times \Theta \\ C^+(E) &= \bigcap_{n=1,2,\dots} \left[B_*^+\right]^n(E) \end{aligned}$$

Action "not invest" is always belief-free rationalizable. Action "invest" is belief-free rationalizable for type $t_i$ of player $i$ exactly if he thinks it is possible that there is common possibility that the state is good, i.e., if $t_i \in B_i^+(C^+(E_G))$. Is this a strong or a weak condition? It is weak in the sense that, at each level, only possibility is required. It is strong in the sense that we still need infinite levels (or a fixed point) to support investment.

To further understand this characterization, note that an event is possible for a player exactly if he is certain of (i.e., assigns probability 1 to) its complement. Thus writing $\sim E$ for the complement of event $E$, we have

$$B_i^+(E) = \sim B_i^1(\sim E).$$

[8] See Monderer and Samet (1989) and Kajii and Morris (1997).

Now suppose that we define an operator $B^1_{**}(E)$ corresponding to *distributed certainty* that $E$ is true, i.e., it corresponds to the event that someone is certain that $E$ is true:

$$B^1_{**}(E) = \left\{(t,\theta) \left| t_i \in B^1_i(E) \text{ for some } i \right.\right\}$$

and say that there is *iterated distributed certainty* of event $E$ if there is distributed certainty of $E$, or there is distributed certainty that there is distributed certainty, or a further iterated version of this statement is true. Thus

$$C^1_{**}(E) = \underset{n=1,2,\ldots}{\cup} \left[B^1_{**}\right]^n (E).$$

Now

$$B^+_*(E) = \sim B^1_{**}(\sim E).$$

Thus, by induction,

$$\left(B^+_*\right)^n (E) = \sim \left(B^1_{**}\right)^n (\sim E)$$

and so

$$C^+(E) = \sim C^1_{**}(\sim E).$$

Thus there is common possibility of event $E$ if and only if there is *not* iterated distributed certainty of *not* $E$. Thus invest is belief-free rationalizable if and only if there is not iterated distributed certainty that the state is bad (i.e., $\theta_B$).

We can also characterize belief invariant Bayes correlated equilibria in this game. There is clearly a belief invariant Bayes correlated equilibrium where each player $i$ invests if he $p$-believes that there is common $p$-belief of the good state, i.e., type $t_i$ invests whenever $t_i \in B^p_i(C^p(E_G))$. Action "invest" is not interim correlated rationalizable for player $i$ if $t_i \notin B^p_i(C^p(E_G))$. Thus the "largest" belief invariant Bayes correlated equilibrium has players investing only on the event $C^p(E_G)$. The structure of Bayes correlated equilibria is more subtle in this example; see Bergemann and Morris (2014) for a discussion of the structure of Bayes correlated equilibria in related contexts.

# 4 Two Important Special Cases

## 4.1 Complete Information

If $\Theta$ is a singleton, then interim correlated rationalizability and belief-free rationalizability will both reduce to (complete information) correlated rationalizability (Brandenburger and Dekel (1987)); and belief invariant Bayes correlated equilibrium and Bayes correlated equilibrium reduce to complete information (objective) correlated equilibrium (Aumann (1987)). In this sense, we are looking at natural generalizations of the classical complete information literature results (at least when they are given an "informational robustness" interpretation).

## 4.2 Payoff Type Spaces

Consider the special case where the payoff relevant states have a product structure, i.e.,

$$\Theta = \Theta_1 \times \Theta_2 \times .... \times \Theta_I$$

and each player knows his own "payoff type" $\theta_i \in \Theta_i$, and nothing more. Thus we have $T_i = \Theta_i$. This "naive" or "payoff" type space is a particular example of a type space as used in the preceding analysis. On this space, beliefs will reduce to $\pi_i : \Theta_i \to \Delta(\Theta_{-i})$.

The definition of interim correlated rationalizability reduces as follows. Let $ICR_i^0(\theta_i) = A_i$ and let $ICR_i^{n+1}(\theta_i)$ equal the set of actions for which there exists $\nu_i \in \Delta(A_{-i} \times \Theta_{-i})$ such that

$$\begin{aligned}
&(1)\ \nu_i(a_{-i}, \theta_{-i}) > 0 \Rightarrow \theta_j \in ICR_j^k(\theta_j) \text{ for each } j \neq i \\
&(2)\ \sum_{a_{-i}} \nu_i(a_{-i}, \theta_{-i}) = \pi_i(\theta_{-i}|\theta_i) \text{ for each } \theta \\
&(3)\ a_i \in \arg\max_{a_i'} \sum_{a_{-i}, \theta_{-i}} \nu_i(a_{-i}, \theta_{-i},) \, u_i((a_i', a_{-i}), (\theta_i, \theta_{-i}))
\end{aligned}$$

and let

$$ICR_i(\theta_i) = \bigcap_{n \geq 1} ICR_i^n(\theta_i).$$

In a payoff type environment, allowing correlation between others' types and payoff states makes no difference here, and this version of interim rationalizability has been widely used in (explicit or implicit) payoff type environments (for example, Battigalli and Siniscalchi (2003a) and Dekel and Wolinsky (2003)).

Belief-free rationalizability will now be defined as follows. Let $BFE_i^0(\theta_i) = A_i$ and let $BFE_i^{n+1}(\theta_i)$ equal the set of actions for which there exists $\nu \in \Delta(A_{-i} \times \Theta_{-i})$ such that

$$\begin{aligned}
&(1)\ \nu(a_{-i}, \theta_{-i},) > 0 \Rightarrow a_j \in BFE_j^n(\theta_j) \text{ for each } j \neq i \\
&(2)\ \sum_{a_{-i}} \nu(a_{-i}, \theta_{-i},) > 0 \Rightarrow \pi_i(\theta_{-i}|\theta_i) > 0 \\
&(3)\ a_i \in \arg\max_{a_i'} \sum_{a_{-i}, \theta_{-i}} \nu(a_{-i}, \theta_{-i})\, u_i((a_i', a_{-i}), (\theta_i, \theta_{-i}))
\end{aligned}$$

and let

$$BFE_i(\theta_i) = \bigcap_{n \geq 1} BFE_i^n(\theta_i).$$

A decision rule will now be a mapping $\sigma : \Theta \to A$ and will be obedient if

$$\begin{aligned}
& \sum_{a_{-i}, \theta_{-i}} \pi^*(\theta_i, \theta_{-i})\, \sigma((a_i, a_{-i}) \,|\, (\theta_i, \theta_{-i}))\, u_i((a_i, a_{-i}), (\theta_i, \theta_{-i})) \\
\geq & \sum_{a_{-i}, \theta_{-i}} \pi^*(\theta_i, \theta_{-i})\, \sigma((a_i, a_{-i}) \,|\, (\theta_i, \theta_{-i}))\, u_i((a_i', a_{-i}), (\theta_i, \theta_{-i})).
\end{aligned}$$

for all $i$, $\theta_i \in \Theta_i$, and $a_i, a_i' \in A_i$; and belief invariant if

$$\sigma_i(a_i | (\theta_i, \theta_{-i})) \triangleq \sum_{a_{-i}} \sigma_i((a_i, a_{-i}) \,|\, (\theta_i, \theta_{-i}))$$

is independent of $\theta_{-i}$. In this case, Bayes correlated equilibrium reduces to the Bayesian solution of Forges (1993) and the belief invariant Bayes correlated equilibrium reduces to belief invariant Bayesian solution of Forges (2006) and Lehrer, Rosenberg, and Shmaya (2010).

Within payoff type spaces, we can consider two further restrictions in order to relate belief-free rationalizability to existing approaches:

1. There are *private values* if $u_i((a_i, a_{-i}), (\theta_i, \theta_{-i}))$ is independent of $\theta_{-i}$. Under the private values assumption, the solution concept of belief-free rationalizability is that studied by Chen, Micali, and Pass (2014b); Chen, Micali, and Pass (2014a) develop novel results about robust revenue maximization using this solution concept.

2. The *full (payoff type) support assumption* is satisfied if $\pi_i(\theta_{-i}|\theta_i) > 0$ for all $i$, $\theta_i$ and $\theta_{-i}$. Under the full support assumption, restriction (2) in the definition of belief-free rationalizability becomes redundant. We referred to this as "incomplete information rationalizability" in Bergemann and Morris (2008). This is the solution concept analyzed in

much of our mechanism design work (Bergemann and Morris (2012)). Note that we did not report beliefs over payoff types in our robust mechanism design work, but if we had, they would be irrelevant to our analysis and we were implicitly assuming full support by always allowing any payoff type profile of others to be associated with a given payoff type of a player. We studied Bayes correlated equilibrium in this context in Bergemann and Morris (2008) where we called it "incomplete information correlated equilibrium".

# 5 Informational Robustness Foundations of Four Solution Concepts

Now suppose that we start out with type space $\mathcal{T}$ and we allow each player $i$ to observe an additional signal $s_i \in S_i$. Each player $i$ has a subjective belief $\phi_i$ about the distribution of signals conditional on the type profiles and the payoff state:

$$\phi_i : T \times \Theta \to \Delta(S).$$

## 5.1 Subjective Belief and the Support Assumption

We make the *support assumption* that, for all players $i$ and $t_i \in T_i$, there exists $\overline{S}_i(t_i) \subseteq S_i$ such that

$$\sum_{s_{-i},t_{-i},\theta} \phi_i\left((s_i, s_{-i}) \,|\, (t_i, t_{-i}), \theta\right) \pi_i\left(t_{-i}, \theta | t_i\right) > 0 \tag{1}$$

for each $s_i \in \overline{S}_i(t_i)$ and

$$\phi_j\left((s_i, s_{-i}) \,| t, \theta\right) = 0 \tag{2}$$

for all $j \neq i$, $s_i \notin \overline{S}_i(t_i)$, $s_{-i}$, $t$ and $\theta$. The interpretation is that if player $i$ has type $t_i$, there is common certainty that he will observe an additional signal $s_i \in \overline{S}_i(t_i)$ and player $i$ thinks that every signal in $\overline{S}_i(t_i)$ is possible. This support assumption ensures that whenever a player other than $i$ thinks that $(t_i, s_i)$ is possible, the beliefs of player $i$ conditional on $(t_i, s_i)$ are well-defined by Bayes rule. This assumption was implicit in the formulation of a correlating device in Liu (2014). We briefly discuss in section 5.2 alternative ways of addressing this issue. We refer to any $(S_i, \phi_i)_{i=1}^{I}$ satisfying the support restriction as an *expansion* of type space $\mathcal{T}$. An expansion is *payoff-irrelevant* if, for each player $i$,

$$\sum_{s_{-i}\in S_{-i}} \phi_i\left((s_i, s_{-i}) \mid (t_i, t_{-i}), \theta\right) \tag{3}$$

is independent of $(t_{-i}, \theta)$. Liu (2014) has shown that this definition characterizes payoff irrelevance in the sense that players can observe signals without altering their beliefs and higher-order beliefs about the state (see also Bergemann and Morris (2014)). Now a basic game $G$, a type space $\mathcal{T}$ and an expansion $(S_i, \phi_i)_{i=1}^{I}$ jointly define a game of incomplete information. A (pure) strategy for player $i$ in this game of incomplete information is a mapping $\beta_i : T_i \times S_i \to A_i$.[9] Now strategy profile $\beta$ is a (Bayes Nash) equilibrium if, for each player $i$, $t_i$ and $s_i \in \overline{S}_i(t_i)$, we have

$$\begin{aligned} &\sum_{t_{-i},s_{-i},\theta} \pi_i(t_{-i}, \theta | t_i)\, \phi_i\left(s_i, s_{-i} | \left((t_i, t_{-i}), \theta\right)\right) u_i\left(\left(\beta_i(t_i, s_i), \beta_{-i}(t_{-i}, s_{-i})\right), \theta\right) \\ \geq &\sum_{t_{-i},s_{-i},\theta} \pi_i(t_{-i}, \theta | t_i)\, \phi_i\left(s_i, s_{-i} | \left((t_i, t_{-i}), \theta\right)\right) u_i\left(\left(a_i, \beta_{-i}(t_{-i}, s_{-i})\right), \theta\right) \end{aligned} \tag{4}$$

for all $a_i \in A_i$.

Now we have informational robustness foundations for the two rationalizability solution concepts we discussed:

**Proposition 1** *Action $a_i$ is interim correlated rationalizable for type $t_i$ of player $i$ in $(G, \mathcal{T})$ if and only if there exists a payoff-irrelevant expansion $\left(S_j, \phi_j\right)_{j=1}^{I}$ of $\mathcal{T}$, an equilibrium $\beta$ of $\left(G, \mathcal{T}, \left(S_j, \phi_j\right)_{j=1}^{I}\right)$ and a signal $s_i \in \overline{S}_i(t_i)$ such that $\beta_i(t_i, s_i) = a_i$.*

Versions of this observation appear as Proposition 2 in Dekel, Fudenberg, and Morris (2006) and as Lemma 2 in Liu (2014). For completeness, and for comparison with the next Proposition, we report a proof in the Appendix for the Proposition under the current notation and interpretation.

**Proposition 2** *Action $a_i$ is belief-free rationalizable for type $t_i$ of player $i$ in $(G, \mathcal{T})$ if and only if there exists an expansion $\left(S_j, \phi_j\right)_{j=1}^{I}$ of $\mathcal{T}$, an equilibrium $\beta$ of $\left(G, \mathcal{T}, \left(S_j, \phi_j\right)_{j=1}^{I}\right)$ and signal $s_i \in \overline{S}_i(t_i)$ such that $\beta_i(t_i, s_i) = a_i$.*

[9] It is without loss of generality to focus on pure strategy profiles for our results: if mixed strategies were allowed, they could always be purified with a richer expansion.

**Proof.** Suppose that action $a_i$ is belief-free rationalizable for type $t_i$ in $(G, \mathcal{T})$. By the definition of belief-free rationalizability, there exists, for each $a_j \in BFR_j(t_j)$, a conjecture $\nu_j^{a_j,t_j} \in \Delta(T_{-j} \times A_{-j} \times \Theta)$ such that

$$
\begin{aligned}
&(1)\ \nu_j^{a_j,t_j}(t_{-j}, a_{-j}, \theta) > 0 \Rightarrow a_k \in BFR_k(t_k) \text{ for each } k \neq j;\\
&(2)\ \sum_{a_{-j}} \nu_j^{a_j,t_j}(t_{-j}, a_{-j}, \theta) > 0 \Rightarrow \pi_j(t_{-j}, \theta | t_j) > 0 \text{ for each } t_{-j}, \theta; \text{ and}\\
&(3)\ a_j \in \arg\max_{a_j'} \sum_{t_{-j}, a_{-j}, \theta} \nu_j^{a_j,t_j}(t_{-j}, a_{-j}, \theta)\, u_j\left(\left(a_j', a_{-j}\right), \theta\right).
\end{aligned}
$$

Now consider the expansion $\left(S_j, \phi_j\right)_{j=1}^I$ of $\mathcal{T}$, where $S_j = A_j \cup \left\{s_j^*\right\}$ and $\phi_j : T \times \Theta \to \Delta(S)$ is given by

$$
\phi_j\left((s_j, s_{-j}) \,|\, (t_j, t_{-j}), \theta\right) = \begin{cases} \frac{\varepsilon}{\#BFR_j(t_j)} \nu_j^{s_j,t_j}(t_{-j}, s_{-j}, \theta), \text{ if } s_j \in BFR_j(t_j) \text{ and } s_{-j} \in BFR_{-j}(t_{-j}) \\ \pi_j(t_{-j}, \theta | t_j) - \varepsilon \sum\limits_{s_{-j} \in A_{-j}} \nu_j^{s_j,t_j}(t_{-j}, s_{-j}, \theta), \text{ if } s_j = s_j^* \text{ and } s_{-j} = s_{-j}^* \\ 0, \text{ otherwise} \end{cases}
$$

for some $\varepsilon > 0$. It is always possible to construct such an expansion for sufficiently small $\varepsilon > 0$ because of property (2) above. Now, by construction, there is an equilibrium of the game $\left(G, \mathcal{T}, \left(S_j, \phi_j\right)_{j=1}^I\right)$ where if $s_j \in \overline{S}_j(t_j)$, $\beta_j(t_j, s_j) = s_j$, and $\beta_j\left(t_j, s_j^*\right)$ can be arbitrarily set equal to any element of

$$
\arg\max_{a_j'} \sum_{t_{-j}, a_{-j}} \pi_j(t_{-j}, \theta | t_j)\, \phi_j\left(s_j^*, a_{-j} | \left((t_j, t_{-j}), \theta\right)\right) u_j\left(\left(a_j', a_{-j}\right), \theta\right)
$$

For the converse, suppose that there exists an expansion $\left(S_j, \phi_j\right)_{j=1}^I$ of $\mathcal{T}$ and an equilibrium $\beta$ of $\left(G, \mathcal{T}, \left(S_j, \phi_j\right)_{j=1}^I\right)$. We will show inductively in $n$ that, for all players $j$, $a_j \in BFR_j^n(t_j)$ whenever $s_j \in \overline{S}_j(t_j)$ and $\beta_j(t_j, s_j) = a_j$. It is true by construction for $n = 0$. Suppose that it is true for $n$. Since $s_j \in \overline{S}_j(t_j)$, equilibrium condition (4) implies that $a_j$ is a best response to a conjecture over others' types and actions and the state. By the inductive hypothesis, this conjecture assigns zero probability to type action profiles $(t_j, a_j)$ of player $j$ where $a_j \notin BFR_j^n(t_j)$. By construction, the marginal of this conjecture on $T_{-j} \times \Theta$ has support contained in the support of $\pi_j(\cdot | t_j)$. Thus $a_j \in BFR_j^{n+1}(t_j)$. ■

An expansion $(S_i, \phi_i)_{i=1}^I$ satisfies the common prior assumption if $\phi_i$ is independent of $i$. An expanded game $\left(G, \mathcal{T}, (S_i, \phi_i)_{i=1}^I\right)$ and a strategy profile $\beta$ for that game will *induce* a decision rule $\sigma : T \times \Theta \rightarrow \Delta(A)$:

$$\sigma(a|t,\theta) = \sum_{\{(t,s):\beta(t,s)=a\}} \phi(s|(t,\theta)).$$

**Proposition 3** *If $\mathcal{T}$ is a common prior type space, then $\sigma$ is a belief invariant Bayes correlated equilibrium of $(G,\mathcal{T})$ if and only if there exists a payoff-irrelevant common prior expansion $(S_i,\phi_i)_{i=1}^I$ of $\mathcal{T}$ and equilibrium $\beta$ of $\left(G,\mathcal{T},(S_i,\phi_i)_{i=1}^I\right)$ such that $\beta$ induces $\sigma$.*

**Proposition 4** *If $\mathcal{T}$ is a common prior type space, then $\sigma$ is a Bayes correlated equilibrium of $(G,\mathcal{T})$ if and only if there exists a common prior expansion $(S_i,\phi_i)_{i=1}^I$ of $\mathcal{T}$ and equilibrium $\beta$ of $\left(G,\mathcal{T},(S_i,\phi_i)_{i=1}^I\right)$ such that $\beta$ induces $\sigma$.*

A subjective version of Proposition 3 appears in Liu (2014) (and the common prior case is discussed in section 5.3). Proposition 4 appears as Theorem 2 in Bergemann and Morris (2014).

## 5.2 The Support Assumption and a Posteriori Equilibrium

In our definitions of expansions, we maintained the *support assumption* (see equations (1) and (2) above). This assumption is automatically satisfied under the common prior assumption. But what is the significance of the support assumption for rationalizability results? If we simply dropped the assumption, and did not replace it with any restriction, then any action could be played in an equilibrium on some expansion of the type space, since it does not matter for ex ante utility what action is played on a zero probability event. An intermediate assumption would be to drop the support assumption, but to add to equilibrium condition (4) the requirement that an action was a best response to some conjecture at signals that were assigned probability zero, i.e., if

$$\sum_{s_{-i},t_{-i},\theta} \phi_i\left((s_i,s_{-i})\,|\,(t_i,t_{-i}),\theta\right)\pi_i(t_{-i},\theta|t_i) = 0, \tag{5}$$

then there exists $\nu_i \in \Delta(T_{-i}\times S_{-i}\times\Theta)$ such that

$$\begin{aligned} &\sum_{t_{-i},s_{-i},\theta} \nu_i(t_{-i},s_{-i},\theta)\,u_i\left(\left(\beta_i(t_i,s_i),\beta_{-i}(t_{-i},s_{-i})\right),\theta\right) \\ \geq &\sum_{t_{-i},s_{-i},\theta} \nu_i(t_{-i},s_{-i},\theta)\,u_i\left(\left(a_i,\beta_{-i}(t_{-i},s_{-i})\right),\theta\right) \end{aligned} \tag{6}$$

for all $a_i \in A_i$. But if we relaxed the support assumption to this intermediate assumption, the resulting solution concept would no longer depend on the type space. In particular, say that an action is *ex post rationalizable* in basic game $\mathcal{G}$ if it survives an iterative deletion procedure where, at each round, we delete actions which are not a best response given any conjecture over surviving actions. Formally, let $EPR_i^0 = A_i$, let $EPR_i^{n+1}$ be the set of actions for which there exists $\nu_i \in \Delta(A_{-i} \times \Theta)$ s.t.

$$\begin{aligned}&(1)\ \nu_i(a_{-i}, \theta) > 0 \Rightarrow a_j \in EPR_j^n \text{ for each } j \neq i,\\ &(2)\ a_i \in \arg\max_{a_i'} \sum_{a_{-i},\theta} \nu_i(a_{-i}, \theta)\, u_i((a_i', a_{-i}), \theta);\end{aligned}$$

and let

$$EPR_i = \bigcap_{n \geq 1} EPR_i^n.$$

So an action could be played in an expanded type space (whether or not (3) is satisfied) if and only if it is ex post rationalizable. Because the solution concept no longer depends on the type space, we would have some counter-intuitive implications. Consider a payoff type space with private values, so that a player has a payoff type determining his private value on the original type space. But because he observes a zero probability signal, we must allow him to have any belief after observing unexpected signals, so he can conclude that his prior belief about his private value was wrong. Our support assumption exactly rules out the possibility of players observing zero probability signals and updating to beliefs outside their original support.

Interim correlated rationalizability, belief-free rationalizability and ex post rationalizability *all* reduce to correlated rationalizability in complete information games. In fact, in the complete information case, allowing players to assign zero probability to signals but requiring them to play best responses to some conjecture is exactly the refinement of subjective correlated equilibrium, *a posteriori equilibrium*, introduced in Aumann (1974) and used in Brandenburger and Dekel (1987). If we wanted to stay closer to the language of Aumann (1974) and Brandenburger and Dekel (1987), we could define a generalized, incomplete information, version of a posteriori equilibrium. If one imposed no new restrictions given the initial type space, we would characterize ex post rationalizability; by imposing additional restrictions at zero probability events, we would get back to belief-free rationalizability and interim correlated rationalizability.

# 6 Epistemic Foundations of Four Solution Concepts

The purpose of this section is twofold. First, we sketch how the "informational robustness" Propositions 1 through 4 of the previous section can be mechanically re-interpreted as classic epistemic results in the spirit of the classic works of Aumann (1987) and Brandenburger and Dekel (1987). Then, we discuss the limitations of these crude translations of our results in the light of the modern epistemic foundations literature. Our purpose in doing so is not to provide novel epistemic foundations results but rather to see how some solution concepts that we and others have worked with in economic applications, and their informational robustness foundations, relate to the epistemic foundations literature.

## 6.1 Epistemic Type Spaces

We fix a basic game $\mathcal{G} = (A_i, u_i)_{i=1}^I$ which will implicitly be assumed to be commonly certain among the players. Let each player have a finite set of types $\Omega_i$. The state space is thus $\Omega \times \Theta$. A type $\omega_i \in \Omega_i$ includes a description of the player's action, given by $\alpha_i : \Omega_i \to A_i$. Let $T_i$ be a finite set and let $\tau_i : \Omega_i \to T_i$. The interpretation is that $t_i \in T_i$ is an arbitrary *partial description* of player $i$'s type. Each player has a belief $\psi_i : T_i \to \Delta(\Omega \times \Theta)$ satisfying

$$\sum_{\omega_{-i},\theta} \psi_i\left(\left(\left(\omega_i, \omega_{-i}\right), \theta\right) | t_i\right) = 0$$

whenever $\tau_i(\omega_i) \neq t_i$ and

$$\sum_{\omega_{-i},\theta} \psi_i\left(\left(\left(\omega_i, \omega_{-i}\right), \theta\right) | t_i\right) > 0$$

whenever $\tau_i(\omega_i) = t_i$. We will refer to $\mathcal{E} = (\Omega_i, \alpha_i, T_i, \tau_i, \psi_i)_{i=1}^I$ as an *epistemic type space.*

For a given epistemic type space, we describe some objects of interest. Let $\widehat{\psi}_i : \Omega_i \to \Delta(\Omega_{-i} \times \Theta)$ be the induced belief of type $\omega_i$ of player $i$ over other players' types and the state:

$$\widehat{\psi}_i(\omega_{-i}, \theta | \omega_i) = \frac{\psi_i\left(\left(\left(\omega_i, \omega_{-i}\right), \theta\right) | \tau_i(\omega_i)\right)}{\sum\limits_{\omega'_{-i},\theta'} \psi_i\left(\left(\left(\omega_i, \omega'_{-i}\right), \theta'\right) | \tau_i(\omega_i)\right)}.$$

An epistemic type space also induces beliefs about partial descriptions of others' types, $\pi_i : T_i \to \Delta(T_{-i} \times \Theta)$, where

$$\pi_i(t_{-i}, \theta | t_i) = \sum_{\{\omega | \tau(\omega) = t\},\theta} \psi_i\left(\left(\left(\omega_i, \omega_{-i}\right), \theta\right) | t_i\right).$$

For an event $E \subseteq \Omega \times \Theta$, let $C_i(E)$ be the set of player $i$'s types who are certain of event $E$, so

$$C_i(E) = \left\{ \omega_i \in \Omega_i \left| \widehat{\psi}_i(\omega_{-i}, \theta | \omega_i) > 0 \Rightarrow ((\omega_i, \omega_{-i}), \theta) \in E \right. \right\}.$$

Let $C_*(E)$ be the event where all players are certain of $E$, so

$$C(E) = \{(\omega, \theta) \in \Omega \times \Theta \,| \omega_i \in C_i(E) \text{ for each } i\}.$$

Let $CC(E)$ corresponding to common certainty of $E$:

$$CC(E) = \bigcap_{n \geq 1} C^n(E).$$

Let $\mathbf{Rat}_i$ be the set of player $i$'s types that are rational:

$$\mathbf{Rat}_i = \left\{ \omega_i \in \Omega_i \left| \alpha_i(\omega_i) \in \arg\max_{a_i} \sum_{\omega_{-i}, \theta} \widehat{\psi}_i(\omega_{-i}, \theta | \omega_i) \, u_i((a_i, \alpha_{-i}(\omega_{-i})), \theta) \right. \right\}.$$

Let $\mathbf{Rat}$ be the event where all players are rational, so

$$\mathbf{Rat} = \{(\omega, \theta) \in \Omega \times \Theta | \omega_i \in \mathbf{Rat}_i \text{ for each } i\}.$$

We now want to discuss players' beliefs and higher-order beliefs about the states $\Theta$. We omit the - by now, standard - construction of this space, following Mertens and Zamir (1985) and Brandenburger and Dekel (1993). This (infinite) space $T^{**}$ satisfies the homeomorphism $T^{**} = \Delta\left((T^{**})^{I-1} \times \Theta\right)$. Now there is a bijection between $T^{**}$ and the set of all beliefs and higher-order beliefs about $\Theta$. We are restricting attention to finite type spaces. Thus we write $T^*$ for the set of types in $T^{**}$ that belong to a belief closed subset of $T^{**}$. We refer to an element of $t^* \in T^*$ as a $\Theta$-*type*, since it is a canonical description of the beliefs and higher-order beliefs of a finite type about $\Theta$.

Now any type $\omega_i \in \Omega_i$ implicitly defines beliefs and higher-order beliefs about $\Theta$. Let the mapping $\gamma_i : \Omega_i \to T^*$ describe the beliefs and higher-order beliefs of player $i$. We will say that $\gamma_i(\omega_i)$ is the $\Theta$-*type* of type $\omega_i$. We also defined a partial description of a player's type, $t_i \in T_i$ and an associated belief about other players' partial types and states. Now partial type $t_i \in T_i$ also implicitly defines beliefs and higher-order beliefs about $\Theta$, specifying beliefs about $\Theta$ conditional on type $t_i$ alone, beliefs about beliefs of others conditional on $T_{-i}$ alone, and $\Theta$, conditional on $t_i$; and so on. Let the mapping $\xi_i : T_i \to T^*$ describe the beliefs and higher-order beliefs of player $i$ based on partial information. We will say that $\xi_i(\tau_i(\omega_i))$ is the partial $\Theta$-*type* of type $\omega_i$.

Finally, recall that we defined solution concepts on arbitrary types spaces. But each of the four solution concepts we have discussed have the property that they are independent of redundant types. Thus if we write $\mathcal{T}^*$ for the type space consisting of $I$ copies of $T^*$ and beliefs $\pi^* : T^* \to \Delta\left((T^*)^{I-1} \times \Theta\right)$, the set of belief-free rationalizable and interim correlated rationalizable actions of a $\Theta$-type of player $i$ in game $(\mathcal{G}, \mathcal{T}^*)$ are well-defined. And the set of BIBCE and BCE decision rules defined on finite belief-closed subsets of $(T^*)^I \times \Theta$ are also well defined.

Now we have the following re-statements of the four informational robustness Propositions above:

**Proposition 5** *If there is common certainty of rationality, then all players are choosing interim correlated rationalizable actions for their $\Theta$-type; i.e.,*

$$(\omega, \theta) \in CC(\mathbf{Rat}) \Rightarrow \alpha_i(\omega_i) \in ICR_i(\gamma_i(\omega_i)).$$

*Conversely, if an action is interim correlated rationalizable for a $\Theta$-type of player $i$, then there exists an epistemic type space and a type of player $i$ in that epistemic type space who takes that action, has that $\Theta$-type and is certain that there is common certainty of rationality; i.e.,*

$$a_i \in ICR_i(t^*) \Rightarrow \exists \mathcal{E} \text{ and } \omega_i \in \Omega_i \text{ such that } \alpha_i(\omega_i) = a_i,\ \gamma_i(\omega_i) = t^* \text{ and } \omega_i \in C_i(CC(\mathbf{Rat})).$$

Various "epistemic foundations" of interim correlated equilibrium have been presented in the literature (Dekel, Fudenberg, and Morris (2006), Battigalli, Di Tillio, Grillo, and Penta (2011) and Liu (2014)). The above statement is closest to that of Theorem 11 of Dekel and Siniscalchi (2014).

**Proposition 6** *If there is common certainty of rationality, then all players are choosing belief-free rationalizable actions for their partial $\Theta$-type; i.e.,*

$$(\omega, \theta) \in CC(\mathbf{Rat}) \Rightarrow \alpha_i(\omega_i) \in BFR_i(\xi_i(\tau_i(\omega_i))).$$

*Conversely, if an action is belief-free rationalizable for a $\Theta$-type of player $i$, then there exists an epistemic type space and a type of player $i$ in that epistemic type space who takes that action, has that partial $\Theta$-type and is certain that there is common certainty of rationality; i.e.,*

$$a_i \in BFE_i(t^*) \Rightarrow \exists \mathcal{E} \text{ and } \omega_i \in \Omega_i \text{ such that } \alpha_i(\omega_i) = a_i,\ \xi_i(\tau_i(\omega_i)) = t^* \text{ and } \omega_i \in C_i(CC(\mathbf{Rat})).$$

An epistemic type space $\mathcal{E} = (\Omega_i, \alpha_i, T_i, \tau_i, \psi_i)_{i=1}^{I}$ satisfies the common prior assumption if there exists a common prior

$$\psi^* \in \Delta(\Omega \times \Theta)$$

such that

$$\sum_{\omega'_{-i},\theta'} \pi^* \left( \left(\omega_i, \omega'_{-i}\right), \theta' \right) > 0$$

for all $i$ and $\omega_i$, and

$$\psi_i \left( \left(\omega_i, \omega_{-i}\right), \theta \right) = 0$$

$$\psi_i \left( \left( \left(\omega_i, \omega_{-i}\right), \theta \right) | \tau_i \left(\omega_i\right) \right) = \frac{\psi^* \left( \left(\omega_i, \omega_{-i}\right), \theta \right)}{\sum\limits_{\left\{\omega'_i | \tau_i\left(\omega'_i\right) = \tau_i(\omega_i)\right\}, \omega'_{-i}, \theta'} \psi_i \left( \left(\omega'_i, \omega'_{-i}\right), \theta' \right)}$$

for all $i$, $(\omega_i, \omega_{-i})$ and $\theta$.

**Proposition 7** *For any common prior epistemic type space, the implied decision rule as a function of $\Theta$-types and states, conditional on common certainty of rationality, is a belief invariant Bayes correlated equilibrium. Conversely, for any belief invariant Bayes correlated equilibrium, there exists a common prior epistemic type space where there is common certainty of rationality and the implied decision rule as a function of $\Theta$-types and states, conditional on common certainty of rationality equals that belief invariant Bayes correlated equilibrium.*

**Proposition 8** *For any common prior epistemic type space, the implied decision rule as a function of $\Theta$-types and states, conditional on common certainty of rationality, is a Bayes correlated equilibrium. Conversely, for any Bayes correlated equilibrium, there exists a common prior epistemic type space where there is common certainty of rationality and the implied decision rule as a function of partial $\Theta$-types and states, conditional on common certainty of rationality, equals that Bayes correlated equilibrium.*

Propositions 5 through 8 are re-writings of Propositions 1 through 4, with (i) different interpretations of the underlying objects; and (ii) restricting attention to type spaces without redundant types.

## 6.2 Discussion

Let us discuss a few of the ways in which these results are "old fashioned" and/or problematic.

1. As in the classical literature, we allowed players to hold beliefs about their own types, i.e., their actions and their rationality. Following Aumann and Brandenburger (1995), we could remove references to agents' beliefs over their own types. It would certainly imply changes in language and, in some cases, such as in the incomplete information extensions of the correlated equilibrium results of Dekel and Siniscalchi (2014), this would lead to novel conceptual issues.

2. We assumed that player $i$ has a belief conditional on his partial type $t_i$ and that, conditional on type $t_i$ alone, he assigns strictly positive probability to every $\omega_i$ with $\tau_i(\omega_i) = t_i$. What does this belief mean? One interpretation is that it is a counterfactual belief: what player $i$ would believe if he only remembered $\tau_i(\omega_i) = t_i$ and did not remember his true type. For most of our analysis, this assumption does not matter and we could have been working directly with $\widehat{\psi}_i : \Omega_i \rightarrow \Delta(\Omega_{-i} \times \Theta)$. But the assumption does matter for Proposition 6. We could work with alternative assumptions. We need to somehow associate a partial type $t_i$ with a hierarchical belief type, but (for purposes of Proposition 6) we could define $\pi_i : T_i \rightarrow \Delta(T_{-i} \times \Theta)$ arbitrarily as long as $\psi_i(\omega_{-i}, \theta|\omega_i) \Rightarrow \psi_i(\tau_{-i}(\omega_{-i}), \theta|\tau_i(\omega_i)) > 0$. But Proposition 6 is perhaps the weakest epistemic foundations result to the extent that we are not comfortable with the interpretation of the belief $\psi_i : T_i \rightarrow \Delta(\Omega \times \Theta)$.

3. We did not impose structure on the epistemic type space and thus allowed distinctions between states which did not correspond to well defined statements in the model about beliefs and higher-order beliefs about exogenous and endogenous variables. Recent works have argued that we should not allow such distinctions; thus Brandenburger and Friedenberg (2008) argue that "extrinsic uncertainty" should not be allowed and Battigalli, Di Tillio, Grillo, and Penta (2011) argue that only events corresponding to "expressible" statements should be allowed. Brandenburger and Friedenberg (2008) argue that - in the context of complete information games - excluding extrinsic uncertainty leads to a slight refinement of correlated rationalizability as the characterization of the implications of common certainty of rationality. Presumably, their approach could be extended to incomplete information and this would similarly refine interim correlated rationalizability. Battigalli, Di Tillio,

Grillo, and Penta (2011) instead enrich the language, by adding explicit signals, to make more statements expressible in the language, and provide expressible epistemic foundations for interim correlated rationalizability.

4. Propositions 7 and 8, like Aumann (1987), show that common certainty of rationality implies the play of a correlated equilibrium. But these results prove implications of common certainty of rationality under the common prior. Thus they rely on an ex ante interpretation of the prior. But epistemic models are usual assumed to be about interim beliefs, and since Aumann (1987), we have elegant characterizations of what it means for a particular type to have beliefs and higher-order beliefs consistent with the common prior (Samet (1998) and Feinberg (2000)). Dekel and Siniscalchi (2014) report an alternative epistemic foundation for correlated equilibrium - in the context of incomplete information games - which takes an interim perspective and avoids discussing a player's belief about his own action. One could presumably extend their approach to incomplete information to give epistemic foundations for belief invariant Bayes correlated equilibrium.

# 7 Appendix

**Proof of Proposition 1.** Suppose that action $a_i$ is interim correlated rationalizable for type $t_i$ in $(G, \mathcal{T})$. By the definition of interim correlated rationalizability, there exists, for each player $j$ and $a_j \in ICR_j(t_j)$, a conjecture $\nu_j^{a_j,t_j} \in \Delta(T_{-j} \times A_{-j} \times \Theta)$ s.t.

$$
\begin{aligned}
&(1)\ \nu_j^{a_j,t_j}(t_{-j}, a_{-j}, \theta) > 0 \Rightarrow a_k \in ICR_k(t_k) \ \text{ for each } k \neq j; \\
&(2)\ \sum_{a_{-j}} \nu_j^{a_j,t_j}(t_{-j}, a_{-j}, \theta) = \pi_j(t_{-j}, \theta | t_j) \ \text{ for each } t_{-j}, \theta; \text{ and} \\
&(3)\ a_j \in \arg\max_{a_j'} \sum_{t_{-j}, a_{-j}, \theta} \nu_j^{a_j,t_j}(t_{-j}, a_{-j}, \theta)\, u_j\left(\left(a_j', a_{-j}\right), \theta\right).
\end{aligned}
$$

Now consider the expansion $\left(S_j, \phi_j\right)_{j=1}^I$ of $\mathcal{T}$, where $S_j = A_j$ and $\phi_j : T \times \Theta \to \Delta(S)$ satisfies

$$
\phi_j\left((a_j, a_{-j}) \,|\, (t_j, t_{-j}), \theta\right) = \begin{cases} \frac{\nu_j^{a_j,t_j}(t_{-j}, a_{-j}, \theta)}{\pi_j(t_{-j}, \theta | t_j) \cdot \# ICR_j(t_j)}, & \text{if} \quad a \in ICR(t), \\ 0 & \text{if} \quad \text{otherwise}; \end{cases}
$$

whenever $\pi_j(t_{-j}, \theta | t_j) > 0$. Now, by construction, there is an equilibrium of the game $(G, \mathcal{T}, \left(S_j, \phi_j\right)_{j=1}^I)$ where

$$
\beta_j(t_j, a_j) = a_j
$$

for all $j$, $t_j$ and $a_j \in ICR_j(t_j)$.

For the converse, suppose that there exists an expansion $\left(S_j, \phi_j\right)_{j=1}^I$ of $\mathcal{T}$, an equilibrium $\beta$ of $(G, \mathcal{T}, \left(S_j, \phi_j\right)_{j=1}^I)$. We will show inductively in $n$ that, for all players $j$, $a_j \in ICR_j^n(t_j)$ whenever $\beta_j(t_j, s_j) = a_j$ for some $s_j \in \overline{S}_j(t_j)$. It is true by construction for $n = 0$. Suppose that it is true for $n$. Equilibrium condition (4) implies that $a_j$ is a best response to a conjecture over others' types and actions and the state. By the inductive hypothesis, this conjecture assigns zero probability to type action profile $(t_k, a_k)$ of player $k \neq j$ with $a_k \notin ICR_k^n(t_k)$. By construction, the marginal on $T_{-j} \times \Theta$ is equal to $\pi_j(\cdot | t_j)$. Thus $a_j \in ICR_j^{n+1}(t_j)$. ■

# References

AUMANN, R. (1974): "Subjectivity and Correlation in Randomized Strategies," *Journal of Mathematical Economics*, 1, 67–96.

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

AUMANN, R., AND A. BRANDENBURGER (1995): "Epistemic Foundations for Nash Equilibrium," *Econometrica*, 63, 1161–1180.

BATTIGALLI, P., A. DI TILLIO, E. GRILLO, AND A. PENTA (2011): "Interactive Epistemology and Solution Concepts for Games with Asymmetric Information," *BE Press Journal of Theoretical Economics: Advances*, 1, Article 6.

BATTIGALLI, P., AND M. SINISCALCHI (2003a): "Rationalizable Bidding in First Price Auctions," *Games and Economic Behavior*, 45, 38–72.

BATTIGALLI, P., AND M. SINISCALCHI (2003b): "Rationalization and Incomplete Information," *Advances in Theoretical Economics*, 3, Article 3.

BERGEMANN, D., AND S. MORRIS (2007): "Belief Free Incomplete Information Games," Discussion Paper 1629, Cowles Foundation for Research in Economics, Yale University.

——— (2008): "The Role of the Common Prior in Robust Implementation," *Journal of the European Economic Association Papers and Proceedings*, 6, 551–559.

——— (2012): *Robust Mechanism Design*. World Scientific Publishing, Singapore.

——— (2013): "Robust Predictions in Games with Incomplete Information," *Econometrica*, 81, 1251–1308.

——— (2014): "Bayes Correlated Equilibrium and the Comparison of Information Structures in Games," Cowles Foundation Discussion Paper 1909RR, Yale University and Princeton University.

BRANDENBURGER, A., AND E. DEKEL (1987): "Rationalizability and Correlated Equilibria," *Econometrica*, 55, 1391–1402.

——— (1993): "Hierarchies of Belief and Common Knowledge," *Journal of Economic Theory*, 59, 189–198.

Brandenburger, A., and A. Friedenberg (2008): "Intrinsic Correlation in Games," *Journal of Economic Theory*, 141, 28–67.

Chen, J., S. Micali, and R. Pass (2014a): "Epistemic Implementation," Discussion paper, MIT.

——— (2014b): "Possibilistic Beliefs and Higher-Level Rationality," Discussion paper, MIT.

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

——— (2007): "Interim Correlated Rationalizability," *Theoretical Economics*, 2, 15–40.

Dekel, E., and M. Siniscalchi (2014): "Epistemic Game Theory," Discussion paper, Northwestern University.

Dekel, E., and A. Wolinsky (2003): "Rationalizable Outcomes of Large Independent Private-Value First Price Discrete Auctions," *Games and Economic Behavior*, 43, 175–188.

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

Forges, F. (1993): "Five Legitimate Definitions of Correlated Equilibrium in Games with Incomplete Information," *Theory and Decision*, 35, 277–310.

——— (2006): "Correlated Equilibrium in Games with Incomplete Information Revisited," *Theory and Decision*, 61, 329–344.

Kajii, A., and S. Morris (1997): "Refinements and Higher Order Beliefs: A Unified Survey," Discussion paper, Northwestern University.

Lehrer, E., D. Rosenberg, and E. Shmaya (2010): "Signaling and Mediation in Games with Common Interest," *Games and Economic Behavior*, 68, 670–682.

Liu, Q. (2014): "Correlation and Common Priors in Games with Incomplete Information," *Journal of Economic Theory*, forthcoming.

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

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

Samet, D. (1998): "Iterated Expectations and Common Priors," *Games and Economic Behavior*, 24, 131–141.

Tan, T., and S. Werlang (1988): "The Bayesian Foundations of Solution Concepts of Games," *Journal of Economic Theory*, 45, 370–391.