BELIEF-FREE RATIONALIZABILITY AND INFORMATIONAL ROBUSTNESS

By

Dirk Bergemann and Stephen Morris

December 2016

COWLES FOUNDATION DISCUSSION PAPER NO. 2066

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

http://cowles.yale.edu/

Belief-Free Rationalizability and Informational Robustness*

Dirk Bergemann† Stephen Morris‡

December 16, 2016

Abstract

We propose an incomplete information analogue of rationalizability. An action is said to be belief-free rationalizable if it survives the following iterated deletion process. At each stage, we delete actions for a type of a player that are not a best response to some conjecture that puts weight only on profiles of types of other players and states that that type thinks possible, combined with actions of those types that have survived so far. We describe a number of applications.

This solution concept characterizes the implications of equilibrium when a player is known to have some private information but may have additional information. It thus answers the "informational robustness" question of what can we say about the set of outcomes that may arise in equilibrium of a Bayesian game if players may observe some additional information.

JEL Classification: C72, C79, D82, D83.

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

*We acknowledge financial support from NSF ICES 1215808. An earlier version of the material reported here appeared as a working paper, Bergemann and Morris (2015), under the title "Informational Robustness and Solution Concepts". We are grateful for discussions with Pierpaolo Battigalli, Eddie Dekel, Françoise Forges, Drew Fudenberg, Qingmin Liu and Marciano Siniscalchi.

†Department of Economics, Yale University, New Haven CT 06511, dirk.bergemann@yale.edu.

‡Department of Economics, Princeton University, Princeton NJ 08544, smorris@princeton.edu.

1 Introduction

We propose a definition of incomplete information rationalizability. A player's possibilities are the set of states that he believes possible. Suppose that we fix each player's possibilities, his possibilities about others' possibilities, and so on (his "higher-order possibilities"). An action is *belief-free rationalizable* for a given higher-order possibility type if it survives the following iterated deletion process. A conjecture of a player is a belief about other players' action profiles, their higher-order possibilities and a payoff relevant state. At each stage of an iterated deletion, delete for each type the set of actions that are not a best response to any conjecture that assigns probability zero to (i) states that that type considers impossible; and (ii) action and type combinations of other players that have already been deleted.

This paper makes a number of contributions. First, we propose this new solution concept for incomplete information games. Second, we study the implication of this solution concept in a number of important economic applications. And third, we show that the solution concept captures the idea of informational robustness: what can we say about rational play in a given environment if we know that players have a certain amount of private information - about payoff relevant states and others' private information - but cannot rule out the possibility that players have additional information? The new solution concept also provides a benchmark for a larger literature looking at informational robustness, which we review.

With respect to the economic applications, we first consider the *payoff-type* environments, where each player knows his own *payoff-type* and thinks every payoff-type profile of other players is possible. In this case, belief-free rationalizability has a simpler characterization. In this setting, a player's payoff-type is a sufficient condition for his higher-order possibility type. Now we can iteratively delete actions for each payoff-type that are not a best response to some belief about others' payoff-types and actions that puts zero probability on action - payoff-type pairs that have already been deleted. Players have unique rationalizable actions if their utilities are sufficiently insensitive to others' types. We provide a tight characterization of when there is sufficient insensitivity to obtain a unique outcome, in linear best response games; we also describe how this result extends to a class of games where each player's utility depends on some sufficient statistic of all other players' payoff-types.

We then depart from known payoff-type environments, and consider two-player two-action games where each player is choosing between a risky action and a safe action. The safe action always gives a payoff of zero. The payoff to the risky action depends on the other player's action and the payoff state. The payoff to the risky action when the other player takes the safe action is always negative, and can be interpreted as the cost of taking the risky action. Two important applications, within this class of games, are studied. In coordination games - where if both players take the risky action, the payoffs of the players always have the same sign - we interpret the risky action as "invest". In trading games - where if both players take the risky action, the payoffs of the players have different signs - we interpret the risky action as accepting a trade. The safe action ("don't invest" or "reject trade") is always belief-free rationalizable in these games. We characterize when the risky action ("invest" or "accept trade") is belief-free rationalizable. An event is said to be *commonly possible* for a player if he thinks that the event is possible (i.e., assigns it strictly positive probability), thinks that it is possible that both the event is true and that the other player thinks it is possible; and so on. Invest (the risky action in the coordination game) is belief-free rationalizable for a player if and only if it is a common possibility for that player that the payoff from both players investing is positive. Accepting trade (the risky action in the trading game) is belief-free rationalizable for a player if and only if an analogous iterated statement about possibility is true: (i) each player thinks that it is possible that he gains from trade; (ii) each player thinks it is possible that both he gains from trade and that (i) holds for the other player; and so on.

To understand the informational robustness foundations of belief-free rationalizability, suppose we start with a fully-specified type space, including players' beliefs as well the support of those beliefs (i.e., the set of states and others' types that are thought possible). Now suppose that players started with the information in that type space but were able to observe additional information. What can we say about a player's updated beliefs? One restriction is that that player cannot assign positive probability to something that was not thought possible. But for some subjective interpretation of the signals a player observes, there will be no other restrictions on what updated beliefs might look like. But our definition of belief-free rationalizability exactly captures these assumptions: the support of beliefs is fixed but not the exact probabilities.

We can also use belief-free rationalizability to understand a larger literature on informational robustness. Belief-free rationalizability is permissive because it allows players to observe payoff-relevant information and does not impose the common prior assumption. If one allows only payoff-*irrelevant* information, i.e., correlating devices, but still without imposing the common prior assumption, then we get the solution concept of *interim correlated rationalizability* (Dekel, Fudenberg, and Morris (2007)). If one imposes the common prior assumption, but allow payoff-relevant information, then we get *Bayes correlated equilibrium* (Bergemann and Morris (2016a)).[1] If one imposes both the common prior assumption and payoff- irrelevant information, then one gets the *belief invariant Bayes correlated equilibrium* studied in Liu (2015). The following table now summarizes these relationships between the solution concepts:

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

Under complete information - the solution concepts without the common prior assumption reduce to the standard notion of correlated rationalizability (Brandenburger and Dekel (1987)), while both correlated equilibrium notions (with the common prior assumption) reduce to the standard notion of (objective) correlated equilibrium Aumann (1987).

One contribution of this paper is to then provide a unified description for informational robustness foundations of these solutions concepts. In each case, we characterize what can happen in (Bayes Nash) equilibrium if we allow players to observe additional information as described above. These informational robustness foundations follow Brandenburger and Dekel (1987) and Aumann (1987) in showing *even if* one makes the strong (and perhaps unjustified[2]) assumption of equilibrium, one cannot remove the possibility of rationalizable play or correlated equilibrium distributions being played if payoff-irrelevant signals are observed (not imposing or imposing the common prior assumption, respectively, in the two cases).

The formal statements in Brandenburger and Dekel (1987) and Aumann (1987) have the informational robustness statements described above.[3] However, both papers interpret their results informally as establishing foundations for solution concepts by establishing that they correspond

[1]In recent work, we have argued that Bayes correlated equilibrium is the relevant tool for characterizing (common prior) robust predictions in games as well as information design (Bergemann and Morris (2013) and Bergemann and Morris (2016b)).

[2]There is a potential tension between assuming equilibrium - a solution concept that has correct common beliefs built into it - in environments where the common prior assumption is not satisfied. Thus Dekel, Fudenberg, and Levine (2004) argue that natural learning justifications that would explain equilibrium in an incomplete information setting would also give rise to a learning justification of common prior beliefs.

[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 equals

to the implications of common certainty of rationality,[4] with or without the common prior assumption: imposing common certainty of rationality is formally equivalent to the assumption of equilibrium on the commonly certain component of the type space. The later literature on "epistemic foundations" has developed more formal statements of these results as consequence of common certainty of rationality.[5] In the current paper, we deliberately focus on a narrower informational robustness interpretation of the results both because this is the interpretation that is relevant for our applications and because the modern epistemic foundations literature addresses a wide set of important but subtle issues that are relevant for the epistemic interpretation but moot for our informational robustness interpretation. Desiderata that are important in the modern epistemic foundations literature are therefore not addressed, including (i) the removal of reference to players' beliefs about their own types or counterfactual belief of types (Aumann and Brandenburger (1995)); (ii) restricting attention to state spaces that reflect "expressible" statements about the model (Brandenburger and Friedenberg (2008) and Battigalli, Di Tillio, Grillo, and Penta (2011)); (iii) giving an interim interpretation of the common prior assumption (Dekel and Siniscalchi (2014)). Battigalli and Siniscalchi (2003) introduced the notion of "Δ-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 environments) are special cases of "Δ-rationalizability", where particular restrictions are placed on beliefs about other players' types and states. Belief-free rationalizability could also be given a Δ-rationalizability

---

the set of 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. Aumann (1974) has an explicit informational robustness motivation.

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

formulation, outside of payoff-type environments, where the corresponding type-dependent restriction on beliefs would be on the support of the beliefs only.

There are two important special cases where belief-free rationalizability has already been applied in payoff-type environments. 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 (2015) have proposed what we are calling belief-free rationalizability in this context and used it for novel results on robust revenue maximization. Payoff-type environments without private values were the focus of earlier work of ours on robust mechanism design collected in Bergemann and Morris (2012); and we report here translations of our mechanism design results on payoff-type environments to general games.[6] Battigalli, Di Tillio, Grillo, and Penta (2011) studied - and used the name - "belief-free rationalizability" in the context of payoff-type environments. We used Bayes correlated equilibrium (in the special case of payoff-type environments) in Bergemann and Morris (2008). The unified treatment of informational robustness thus also embeds both our earlier work on robust mechanism design and our more recent work on robust predictions in games (Bergemann and Morris (2013), (2016)).

The informational-robustness results in this paper concern what happens if players observe extra signals about payoffs, but without allowing payoff perturbations. A related but different strand of the literature (Fudenberg, Kreps, and Levine (1988), Kajii and Morris (1997) and Weinstein and Yildiz (2007)) examines the robustness of equilibrium predictions to payoff perturbations about which players face uncertainty.

We define the notion of belief-free rationalizability in Section 2. We develop the implications of belief-free rationalizability in a number of applications in Section 3. We relate belief-free rationalizability to three other, previously introduced, solution concepts in Section 4. There we also give unified informational-robustness foundations for all of these solution concepts. In the final Section 5, we discuss the support assumption and relate it the notion of *a posteriori equilibrium* of Aumann (1974) in complete information games.

[6] Our working paper, Bergemann and Morris (2007), covered some of the same material as this paper for payoff-type environments and is thus incorporated in this paper.

2 Setting and Belief Free Rationalizability

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

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 $A = A_1 \times \cdots \times A_I$. 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 player, a belief over others' 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 $\mathcal{T} = (T_i, \pi_i)_{i=1}^I$.

In defining belief-free rationalizability, it is only the support of $\pi_i(t_i)$ that matters, not the (strictly positive) probabilities assigned to elements of that support. We allow the implied redundancies in our description of the type space to facilitate later comparisons with other solution concepts.

We define belief-free rationalizability inductively as follows. Suppose $BFR_i^n(t_i)$ describes the $n-$th level (belief-free) rationalizable actions for type t_i of agent i. Write BFR^n for the profile of correspondences where each BFR_i^n is a non-empty correspondence $BFR_i^n : T_i \to 2^{A_i} / \varnothing$. We will say that a_i is not (belief-free) dominated for type t_i with respect to BFR^n if there exists a conjecture over profiles of other players' actions, types and payoff relevant states, whose support is consistent with BFR^n and the support of that type's beliefs. Thus writing $BFR_i^n(t_i)$ for the set of actions that survive n rounds of deletion, we let $BFR_i^0(t_i) = A_i$, let $BFR_i^{n+1}(t_i)$ be the set of actions for which there exists a conjecture $\nu_i \in \Delta(T_{-i} \times A_{-i} \times \Theta)$ such that

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

and let

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

Definition 1 (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 numerical beliefs and depends only on which profiles of other players' types and states he considers possible, i.e., the support of his beliefs.

3 Applications of Belief-Free Rationalizability

In this section, we investigate the implications of belief-free rationalizability in some well-known economic environments: first, linear best response games and, second, coordination and trading games.

3.1 Payoff-Type Environments and Linear Best-Response Games

We first study *payoff-type* environments. We suppose that each player i has a *payoff-type* θ_i that he knows; and that the payoff state θ is just the profile of players' payoff-types, $\theta = (\theta_i)_{i=1}^I$. This assumption is maintained in many settings (including throughout our own work on robust mechanism design collected in Bergemann and Morris (2012)). We will also maintain a full support assumption: all types of all players - while knowing their own payoff-types - think that every profile of others' payoff-types are possible (this assumption was implicit in our work on robust mechanism design). Under these assumptions, a payoff-type is a sufficient statistic for a player's higher-order possibilities, since every type is certain of his own payoff-type and there is common certainty that no player is ever certain of anything else. Thus we can identify types with payoff-types for purposes of defining belief-free rationalizability. The definition of belief-free rationalizability now simplifies. Writing $BFR_i^n(\theta_i)$ for the set of actions that survive n rounds of deletion, we have $BFR_i^0(t_i) = A_i$, $BFR_i^{n+1}(\theta_i)$ equal to the set of actions for which there exists a conjecture $\nu_i \in \Delta(A_{-i} \times \Theta_{-i})$ such that

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

and

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

We illustrate belief-free rationalizability in payoff-type spaces by considering a linear best-response game. Such games arise in a wide variety of settings (and arose endogenously out of a mechanism design problem in our prior work, e.g. Bergemann and Morris (2009)). For this section, we will have continuum - instead of finite - actions and payoff-types. We could formally extend our earlier definitions to such continuum action type settings straightforwardly but at the expense of additional notation and qualifications.

So suppose now that $A_i = \Theta_i = [0,1]$ for all i and that agent i has payoff-type θ_i and belief $\mu_i \in \Delta(A_{-i} \times \Theta_{-i})$. We suppose that agent i has a best response to set his action equal to:

$$a_i = \theta_i + \gamma \mathbb{E}_{\mu_i}\left(\sum_{j \neq i}(a_j - \theta_j)\right). \tag{2}$$

Thus each player wants to set his action equal to the payoff state θ_i but make a linear adjustment based on the distance of others' actions from their payoff-types. If the parameter γ is positive, then this is a game with strategic complementarities, while a negative γ corresponds to a game with strategic substitutes. Many payoffs could give rise to this best response function. In particular, this best response function could arise from a common interest game:

$$\begin{aligned} u_i(a,\theta) &= v(a,\theta) \\ &= -\sum_{j=1}^{I}(a_j - \theta_j)\left[(a_j - \theta_j) + \gamma \sum_{k \neq j}(a_k - \theta_k)\right] \\ &= -\sum_{j=1}^{I}(a_j - \theta_j)^2 - \gamma \sum_{j=1}^{I}(a_j - \theta_j)\sum_{k \neq j}(a_k - \theta_k), \end{aligned}$$

for some $\gamma \in \mathbb{R}$. In this case, a player's utility from choosing action a_i is

$$\mathbb{E}_{\mu_i}(v|a_i,\theta_i) = -\int_{a_{-i},\theta_{-i}}\left(\sum_{j=1}^{I}(a_j - \theta_j)^2 + \gamma \sum_{j=1}^{I}(a_j - \theta_j)\sum_{k \neq j}(a_k - \theta_k)\right) d\mu_i.$$

The first-order condition for this problem is then:

$$\frac{d\mathbb{E}_{\mu_i}(v|a_i,\theta_i)}{da_i} = -2(a_i - \theta_i) - 2\gamma \mathbb{E}_{\mu_i}\left(\sum_{j \neq i}(a_j - \theta_j)\right),$$

and setting this equal to zero gives agent i's best response (2). This game has a unique ex post equilibrium, where each player sets his action equal to his payoff-type. The set of belief-free rationalizable actions are:

$$BFR_i(\theta_i) = \begin{cases} \{\theta_i\}, \text{ if } -\frac{1}{I-1} < \gamma < \frac{1}{I-1}; \\ [0,1], \text{ otherwise.} \end{cases}$$

This can be shown inductively:

$$BFR_i^k(\theta_i) \triangleq \left[\max\left\{0, \theta_i - (|\gamma|(I-1))^k\right\}, \min\left\{1, \theta_i + (|\gamma|(I-1))^k\right\}\right].$$

If $|\gamma| < \frac{1}{I-1}$, the set of kth-level rationalizable actions shrinks at every iteration. If $|\gamma| > \frac{1}{I-1}$, the bounds on the kth-level rationalizable actions explode so no action is excluded. In Bergemann and Morris (2009) we show how this logic can be generalized to asymmetric linear best response games when the best response function of player i is given by:

$$a_i = \theta_i - \mathbb{E}_{\mu_i}\left(\sum_{j \neq i} \gamma_{ij}(a_j - \theta_j)\right);$$

and to general games where a player's best response is monotonic in his payoff state and an aggregate statistic of other players' actions. In both cases, there is a unique belief-free rationalizable action if and only if players' utilities are not too sensitive to other players' payoff-types.

3.2 Binary Actions: Coordination and Trade

We now consider some classic economic problems - coordination and trade. For simplicity, we focus our attention on a class of two-player two-action games where the payoffs in state $\theta \in \Theta$ are given by:

θ	Risky	Safe
Risky	$x_1(\theta) - c, x_2(\theta) - c$	$-c, 0$
Safe	$0, -c$	$0, 0$

, (3)

where the risky payoff $(x_1(\theta), x_2(\theta)) \in \mathbb{R}^2$ depends on the realized payoff-state θ. We will characterize belief-free rationalizable actions in this class of games, with additional restrictions giving coordination and trading interpretations. Before presenting these characterizations, we report a general language for discussing higher-order possibility and common possibility that is useful in the characterization of rationalizable behavior in both classes of games.

3.2.1 Higher-Order and Common Possibility

For a fixed type space $\mathcal{T}$, an event E is a subset of $T \times \Theta$. "Possibility operators" are defined as follows. We present the definitions here for two players and the generalization to many players is immediate, but not necessary for our purpose here. We write $B_i(E)$ for the set of types of player i that think that E is possible:

$$B_i(E) = \left\{ t_i \in T_i \left| \begin{array}{l} \exists t_j \in T_j \text{ and } \theta \in \Theta \text{ such that} \\ ((t_i, t_j), \theta) \in E \text{ and } \pi_i(t_j, \theta | t_i) > 0 \end{array} \right. \right\}.$$

For a pair of events $E_1 \subseteq T_1$ and $E_2 \subseteq T_2$, (E_1, E_2) are a *common possibility* for player i if:

1. player i thinks it is possible that E_i is true,
2. player i thinks it is possible that both (i) E_i is true; and (ii) player j thinks that E_j is possible,
3. and so on... .

Thus if we write $C_i(E_1, E_2)$ for the set of types of player i for whom (E_1, E_2) are a common possibility, we have

$$C_i(E_1, E_2) = B_i(E_i) \cap B_i(E_i \cap B_j(E_j)) \cap B_i(E_i \cap B_j(E_j \cap B_i(E_i))) \cap .....$$

More formally, define operators B_1^k and B_2^k on pairs of events by $B_i^0(E_1, E_2) = T_i$ and $B_i^{k+1}(E_1, E_2) = B_i\left(E_i \cap B_j^k(E_j)\right)$ for each $k = 1, 2...,$ we have

$$C_i(E_1, E_2) = \bigcap_{k \geq 1} B_i^k(E_1, E_2). \tag{4}$$

The sequence $B_i^k(E_1, E_2)$ is decreasing under set inclusion for each i. Thus this definition of common possibility also has a well-defined fixed point characterization:

Lemma 1 (Common Possibility as Fixed Point)
Let $F_1 \subseteq T_1$ and $F_2 \subseteq T_2$ be the largest sets of types satisfying $F_1 \subseteq B_1(E_1 \cap F_2)$ and $F_2 \subseteq B_2(E_2 \cap F_2)$. Then $C_i(E_1, E_2) = F_i$.

The definition given by (4) describes a concept of common possibility for a *pair* of events (E_1, E_2) for the two players. If we are only interested in a single event, and we can adapt the above definitions to a single event $E_1 = E_2 = E$, so that event E is a common possibility for player i if:

1. player i thinks it is possible that E is true,

2. player i thinks it is possible that both (i) E is true; and (ii) player j thinks that E is possible,

3. and so on... .

We will write $C_i(E)$ as shorthand for $C_i(E, E)$, and so

$$C_i(E) = B_i(E) \cap B_i(E \cap B_j(E)) \cap B_i(E \cap B_j(E \cap B_i(E))) \cap .....;$$

and this is equivalent to inductively defining

$$B_i^0(E) = T_i \text{ and } B_i^{k+1}(E) = B_i\left(E \cap B_j^k(E)\right),$$

and setting

$$C_i(E) = \bigcap_{k \geq 1} B_i^k(E).$$

3.2.2 Coordination Games

We now return to the two person two action game described by (3) above. For the class of coordination games, define Θ_G to be the set of "good" payoff states where both players strictly benefit if both take the risky action ("invest"); thus

$$\Theta_G = \{\theta \in \Theta \,|x_1(\theta) > c \text{ and } x_2(\theta) > c\}.$$

Define Θ_B to be the set of "bad" payoff states where both players are strictly made worse off even if both take the risky action; thus

$$\Theta_B = \{\theta \in \Theta \,|x_1(\theta) < c \text{ and } x_2(\theta) < c\}.$$

Here, c has the interpretation that it is a cost of investment that is occurred independent of whether others invest. We will define a coordination game to be a situation where all states are either good or bad, so that

$$\Theta = \Theta_G \cup \Theta_B.$$

To remove uninteresting cases based on indifference, this definition excludes the possibility that $x_i(\theta) = c$. We write E_G and E_B for the set of states where the payoff state is good and bad, respectively, so

$$E_G = \{(t, \theta) \,| \theta \in \Theta_G\} \text{ and } E_B = \{(t, \theta) \,| \theta \in \Theta_B\}.$$

In coordination games, at all good states, the corresponding complete information game has two strict Nash equilibria (both invest and both not invest), while at all bad states, both players have a strictly dominant strategy to not invest. Now we have:

Proposition 1 (Belief-Free Rationalizability in Coordination Game)
In a coordination game, the safe action (not invest) is always belief-free rationalizable; the risky action (invest) is belief-free rationalizable for player i if and only if the event E_G is a common possibility for player i.

The first claim follows immediately because both not invest is always a strict Nash equilibrium of the underlying complete information game. For the second claim, observe that $B_i^k(E_G)$ is the set of types of agent i for whom invest is kth level belief-free rationalizable. This follows by induction since $B_i^0(E_G) = T_i$ corresponds to the set of types for whom invest is 0th level belief-free rationalizable; and, if $B_j^k(E_G)$ is the set of types of player j for whom invest is kth level belief-free rationalizable, then invest is $(k+1)$th level rationalizable for player i only if he attaches positive probability to $E_G \cap B_j^k(E_G)$. But - by definition - the set of types of player i for which this is true is exactly $B_i^{k+1}\left(E_G \cap B_i^k(E_G)\right)$, so we have our induction.

3.2.3 Trading Games

We now want to consider a class of trading games where the safe action is interpreted as no trade and the risky action is interpreted as (agreeing to) trade. For this exercise, we think of c as being very small and corresponding to a small transaction cost associated with agreeing to trade. But trade will only take place if both players agree to trade. Let $\Theta_i \subseteq \Theta$ be the set of "i gain (payoff) states" where trade is beneficial for player i, but not for player j, so

$$\Theta_i = \{\theta \in \Theta \,| x_i(\theta) > c \text{ and } x_j(\theta) < c\}.$$

Define a trading game to be a situation where all states are gain states for exactly one player, so that

$$\Theta = \Theta_1 \cup \Theta_2.$$

True zero-sum trade would require that $x_1(\theta) + x_2(\theta) = 0$ for all θ, while a weaker non-positive sum trade requirement would be that $x_1(\theta) + x_2(\theta) \leq 0$ for all θ. We do not use either of these restrictions for our results and we would not get sharper results if we imposed either of them. Now write E_i^+ for the set of states and types corresponding to i-gain payoff state for player i,

$$E_i^+ = \{(t,\theta) \,|\theta \in \Theta_i\}.$$

Now we have:

Proposition 2 (Belief-Free Rationalizability in Trading Game)
In a trading game, the safe action (reject trade) is always belief-free rationalizable; the risky action (accept trade) is belief-free rationalizable for player i if and only if the events $\left(E_1^+, E_2^+\right)$ are a common possibility for player i.

The first claim is immediate because, in a trading game, the strictly positive cost c implies that there is always a strict equilibrium where each player never trades, which in turn implies that rejecting trade must be belief-free rationalizable. For the second claim, observe that $B_i^k\left(E_i^+, E_j^+\right)$ is the set of types of player i for whom trade is kth level belief-free rationalizable. This follows by induction: $B_i^0\left(E_i^+, E_j^+\right) = T_i$ corresponds to the set of types from whom accepting trade is 0th level belief-free rationalizable; and if $B_j^k\left(E_j^+, E_i^+\right)$ is the set of types of player j for whom accepting trade is k-th level belief-free rationalizable, then trade is $(k+1)$-th level belief-free rationalizable for player i only if he attaches positive probability to $E_i \cap B_j^k\left(E_j^+, E_i^+\right)$. But - by definition - the set of types of player i for which this is true is exactly $B_i^{k+1}\left(E_i^+ \cap B_j^k\left(E_j^+, E_i^+\right)\right)$, so we have our induction.

We follow Morris and Skiadas (2000) in proving a no trade result under rationalizability.[7] While they did not explicitly use belief-free rationalizability as a solution concept, their results are remain true as stated for this definition and our proposition is essentially the same as theirs, although their characterization is expressed in very different language. In the working paper version of this paper Bergemann and Morris (2015), we show how our characterization reduces to the one reported there.

[7] Morris and Skiadas (2000) maintained the payoff type assumption, so that trades were conditional on only the type profile.

4 Solution Concepts and Informational Robustness

We now discuss three more solution concepts in order to put belief-free rationalizability in context and provide unified informational-robustness foundations of solution concepts.

4.1 Three More Solution Concepts

First, consider *interim correlated rationalizability* (Dekel, Fudenberg, and Morris (2007)), which is a refinement of belief-free rationalizability. 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 $ICR_i^0\left(t_i\right)=A_i$ and let $ICR_i^{n+1}\left(t_i\right)$ equal the set of actions for which there exists $\nu_i \in \Delta\left(A_{-i}\times T_{-i}\times\Theta\right)$ such that

$$\begin{aligned}
&(1)\ \nu_i\left(a_{-i},t_{-i},\theta\right)>0 \Rightarrow a_j \in ICR_j^n\left(t_j\right) \text{ for each } j\neq i;\\
&(2)\ \sum_{a_{-i}} \nu_i\left(a_{-i},t_{-i},\theta\right)=\pi_i\left(t_{-i},\theta|t_i\right) \text{ for each } t_{-i},\theta;\\
&(3)\ a_i \in \arg\max_{a_i'} \sum_{a_{-i},t_{-i},\theta} \nu_i\left(a_{-i},t_{-i},\theta\right) u_i\left(\left(a_i',a_{-i}\right),\theta\right);
\end{aligned} \tag{5}$$

and let

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

Definition 2 (Interim Correlated Rationalizable)
Action a_i is interim correlated rationalizable for type t_i (in game $(\mathcal{G},\mathcal{T})$) if $a_i \in ICR_i\left(t_i\right)$.

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

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

for all i and t_i, and

$$\pi_i\left(t_{-i},\theta|t_i\right)=\frac{\pi^*\left(\left(t_i,t_{-i}\right),\theta\right)}{\sum\limits_{t'_{-i},\theta'}\pi^*\left(\left(t_i,t'_{-i}\right),\theta'\right)},$$

for all i, (t_i,t_{-i}) and θ.[8]

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, $\sigma:T\times\Theta\rightarrow\Delta(A)$. A decision rule σ satisfies *belief-invariance* if, for each player i,

$$\sigma_i\left(a_i|\left(t_i,t_{-i}\right),\theta\right)\triangleq\sum_{a_{-i}}\sigma\left(\left(a_i,a_{-i}\right)|\left(t_i,t_{-i}\right),\theta\right) \tag{6}$$

is independent of (t_{-i},θ). Thus a decision rule satisfies belief-invariance if a player's action 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 in (5) was a belief-invariance assumption.

Decision rule σ 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 σ 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.

[8]When the common prior assumption is maintained, we understand the common prior π^* 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 in the literature. This version simplifies the statement of our results and will also tie in with the support assumption that we impose in the informational robustness foundations in Section 4.3.

Definition 3 (Belief Invariant Bayes Correlated Equilibrium (BIBCE))
Decision rule σ is a belief invariant Bayes correlated equilibrium (BIBCE) if it satisfies obedience and belief-invariance.

Liu (2015) described the subjective correlated equilibrium analogue of interim correlated rationalizability. If one then imposes the common prior assumption (as he discusses in Section 5.2), then the version of incomplete information correlated equilibrium that one obtains is given by Definition 3.[9] Its relation to the incomplete information correlated equilibrium literature is further discussed in Bergemann and Morris (2016a): 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 the 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 (BCE))
Decision rule σ is a Bayes correlated equilibrium (BCE) if it satisfies obedience.

This solution concept is studied in Bergemann and Morris (2016a). It is immediate from the definitions that any action played with positive probability by a type in a Bayes correlated equilibrium is belief-free rationalizable.

4.2 Back to the Applications

We will selectively report what happens to some of our earlier applications under these three additional solution concepts.

4.2.1 Payoff-Type Spaces and Linear Best Response Games

The three new solution concepts lead to simpler statements and interpretations in the special case of payoff-type environments. In a payoff-type environment, there is "distributed certainty":

[9] Liu (2015) refers to what we are calling the belief invariant Bayes correlated equilibrium as the "common-prior correlated equilibrium" (see Definition 4 in Liu (2015)). We use the current language to emphasize the belief invariance property relative to the Bayes correlated equilibrium itself.

the join of players' information reveals the true state . But under this assumption, the "correlation" in interim *correlated* 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).

One can show that if

$$\gamma \in \left[-\frac{1}{I-1}, 1\right],$$

there is a unique Bayes correlated equilibrium in the linear best response example. In this equilibrium, each player sets his action equal to his payoff-type. Thus if $1 < \gamma < -\frac{1}{I-1}$, then there is a unique Bayes correlated equilibrium but all actions are belief-free rationalizable. This follows from results in Bergemann and Morris (2008), via a potential game argument.

4.2.2 Binary Actions: Coordination and Trade

Higher-Order and Common Beliefs We now introduce belief operators analogous to the possibility operators introduced earlier. We will use both possibility and p-belief operators to analyze the coordination and trading game under interim correlated rationality and belief-invariant Bayes correlated equilibrium. We write $B_i^p(E)$ for the set of types of player i who assign probability at least p to event E,

$$B_i^p(E) = \left\{ t_i \in T_i \left| \sum_{\{(t_j,\theta) | ((t_i,t_j),\theta) \in E\}} \pi_i(t_j, \theta | t_i) \geq p \right. \right\}.$$

The connection with possibility operators is that if a player assigns probability p to an event for any $p > 0$, then he thinks that the event is possible:

$$B_i(E) = \bigcup_{p>0} B_i^p(E).$$

Event E is *repeated common p-belief* for player i if:

1. player i assigns probability at least p to event E,

2. player i assigns probability at least p to the event that both (i) E is true; and (ii) player j assigns probability at least p to event E,

3. and so on... .

Now writing $C_i^p(E)$ for the set of types of player i for whom event E is *repeated common p-belief*, we have that

$$C_i^p(E) = B_i^p(E) \cap B_i^p\left(E \cap B_j^p(E)\right) \cap B_i^p\left(E \cap B_j^p(E \cap B_i^p(E))\right) \cap .....;$$

and this is equivalent to inductively defining

$$B_i^{p,0}(E) = T_i \text{ and } B_i^{p,k+1}(E) = B_i^p\left(E \cap B_j^{p,k}(E)\right),$$

and setting

$$C_i^p(E) = \underset{k\geq 1}{\cap} B_i^{p,k}(E).$$

This definition of belief operators follows Monderer and Samet (1989) while the definition of repeated common p-belief comes from Monderer and Samet (1996).[10]

Coordination We now focus on a more specialized class of coordination games. Suppose that

$$\begin{aligned} \theta &\in \Theta_G \Rightarrow x_1(\theta) = x_2(\theta) = x > c, \\ \theta &\in \Theta_B \Rightarrow x_1(\theta) = x_2(\theta) = 0. \end{aligned}$$

Call this a *simple* coordination game.

Proposition 3 (Belief-Free Rationalizability in a Simple Coordination Game) *In a simple coordination game, the safe action (not invest) is always interim correlated rationalizable; the risky action (invest) is interim correlated rationalizable if and only if the event E_G is repeated common c/x-belief for player i.*

Again, the first claim follows immediately because since both not invest is always a strict Nash equilibrium of the underlying complete information game. For the second claim, note that

[10]The definition of repeated common p-belief is closely related to the more widely used concept of common p-belief introduced in Monderer and Samet (1989) given by

$$\widetilde{C}_i^p(E) = B_i^p(E) \cap B_i^p(B_1^p(E) \cap B_2^p(E)) \cap B_i^p(B_1^p(B_1^p(E) \cap B_2^p(E)) \cap B_2^p(B_1^p(E) \cap B_2^p(E))) \cap .....;$$

Monderer and Samet (1996) describe the close relationship between common p-belief and repeated common p-belief, which we omit here.

invest is a best response for a player only if he attaches probability at least c/x to both the state being good and his opponent choosing to invest. Now, analogously to belief-free rationalizability case, we can show by induction that $B_i^{\frac{c}{x},k}(E_G)$ is the set of types of agent i for whom invest is $k-$th level belief-free rationalizable: $B_i^{\frac{c}{x},0}(E) = T_i$ is the set of types from whom invest is 0th level belief-free rationalizable, and, if $B_j^{\frac{c}{x},k}(E)$ is the set of types of player j for whom invest is kth level belief-free rationalizable, then invest is $(k+1)$th level rationalizable for player i only if he attaches probability at least c/x to $E \cap B_j^{\frac{c}{x},k}(E)$ and so, again, we have our induction.

Because this is a game of strategic complementarities, and the largest and smallest rationalizable strategies (in the natural order) constitute equilibria, we have:

Proposition 4 (Belief Invariant BCE in a Simple Coordination Game)
In a simple coordination game, there is a belief-invariant Bayes correlated equilibrium where the safe action (not invest) is always played. There is another belief-invariant Bayes correlated equilibrium where the risky action (not invest) is played by player i if and only if the event E_G is repeated common c/x-belief. All other belief-invariant Bayes correlated equilibria are "in between" these two, in the sense that invest is never played if the event E_G is not repeated common c/x-belief.

The structure of Bayes correlated equilibria is more subtle in this example; see Bergemann and Morris (2016a) for a discussion of the structure of Bayes correlated equilibria in two-player two-action games of incomplete information.

Trading The characterization of belief-free rationalizability extends almost immediately to interim correlated rationalizability:

Proposition 5 (Interim Correlated Rationalizability in Trading Game)
In a trading game, the safe action (reject trade) is always interim correlated rationalizable; the risky action (accept trade) is interim correlated rationalizable for player i if and only if the events (E_1, E_2) are a common possibility of player i.

To see why, it is enough to show that the inductive step that worked for belief-free rationalizability continues to work for interim correlated rationalizability. In particular, suppose that E_j^k is the set of types of player j for whom accept trade is k-th level rationalizable (recall that rejecting trade is always kth level rationalizable). Now consider a type t_i of player i. He will have an interim belief $\pi_i(\cdot|t_i)$ over (t_j, θ), the type of the other player and the payoff state. Suppose $\left(t_j^*, \theta^*\right) \in E_j^k \times \Theta_i$, i.e., a type payoff state pair where accept trade is k-th level interim correlated rationalizable for player j and the payoff state is an i-gain state. Now we can endow type t_i of agent with a belief $\nu_i \in \Delta(A_j \times T_j \times \Theta)$ given by

$$\nu_i(a_j, t_j, \theta) = \begin{cases} \pi_i(t_j, \theta|t_i), & \text{if } \quad a_j = \text{reject trade and } (t_j, \theta) \neq \left(t_j^*, \theta^*\right), \\ \pi_i(t_j, \theta|t_i), & \text{if } \quad a_j = \text{accept trade and } (t_j, \theta) = \left(t_j^*, \theta^*\right), \\ 0, & \text{if } \quad \text{otherwise.} \end{cases}$$

Clearly, accept trade is best response to this conjecture and thus $(k+1)$-th level rationalizable for type t_i. Thus the induction argument for belief-free rationalizability goes through unchanged for interim correlated rationalizability.

In the common prior case, we have

Proposition 6 (BCE in Trading Game)
In a trading game, there is a unique Bayes correlated equilibrium (and thus a unique belief invariant Bayes correlated equilibrium) where both players always choose the safe action (reject trade).

It is well known that trade is not possible under the common prior assumption: see Sebenius and Geanakoplos (1983) for a statement in the bilateral risk neutral setting discussed here and Milgrom and Stokey (1982) in a more general environment. Arguments from this literature immediately apply.

4.3 Informational Robustness Foundations

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 ϕ_i about the distribution of signals conditional on the type profiles and the payoff state:

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

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

$$\sum_{s_{-i}, t_{-i}, \theta} \phi_i\left((s_i, s_{-i}) \,|\, (t_i, t_{-i}), \theta\right) \pi_i(t_{-i}, \theta|t_i) > 0, \tag{7}$$

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

$$\phi_j((s_i, s_{-i})|t, \theta) = 0, \tag{8}$$

for all $j \neq i$, $s_i \notin \overline{S}_i(t_i)$, s_{-i}, t and θ. The interpretation is that if player i does not think it is possible that he will observe an additional signal $s_i \notin \overline{S}_i(t_i)$ if he is type t_i, then no player j ever thinks it is possible that player i observes signal s_i when his type is t_i. 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. If this assumption is not made, then players can attach positive probability to other players being types with undefined beliefs. Aumann (1974) discussed why an assumption like this was necessary in a sensible definition of subjective correlated equilibrium with an informational robustness interpretation. This assumption was implicit in the formulation of a correlating device in Liu (2015). We briefly discuss in Section 5 alternative ways of addressing this issue and the relation to "a posteriori equilibrium" in the complete information case.

We refer to any conditional distribution of signals, $(S_i, \phi_i)_{i=1}^I$, satisfying the support restriction as an *expansion* of type space $\mathcal{T}$. An expansion is *belief-invariant* if, for each player i,

$$\sum_{s_{-i} \in S_{-i}} \phi_i((s_i, s_{-i}) \mid (t_i, t_{-i}), \theta), \tag{9}$$

is independent of (t_{-i}, θ). Note that this is the same definition as (6) applied to expansions rather than decision rules, and it will immediately translate into belief-invariance of decision rules in our informational robustness results. Liu (2015) 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 (2016a)).

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 $\beta_i : T_i \times S_i \to A_i$. Now strategy profile β is a (Bayes Nash) equilibrium if, for each player i, t_i and $s_i \in \overline{S}_i(t_i)$, we have

$$\begin{aligned} & \sum_{t_{-i}, s_{-i}, \theta} \pi_i(t_{-i}, \theta|t_i)\, \phi_i(s_i, s_{-i}| ((t_i, t_{-i}), \theta))\, u_i\left(\left(\beta_i(t_i, s_i), \beta_{-i}(t_{-i}, s_{-i})\right), \theta\right) \\ \geq & \sum_{t_{-i}, s_{-i}, \theta} \pi_i(t_{-i}, \theta|t_i)\, \phi_i(s_i, s_{-i}| ((t_i, t_{-i}), \theta))\, u_i\left(\left(a_i, \beta_{-i}(t_{-i}, s_{-i})\right), \theta\right), \end{aligned} \tag{10}$$

for all $a_i \in A_i$.

Now we can formally state the informational robustness foundations for the two rationalizability solution concepts we discussed:

Proposition 7 (Informational Robustness to Payoff-Irrelevant Signals)
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 β 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 (2007) and as Lemma 2 in Liu (2015). 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 8 (Informational Robustness to Payoff-Relevant Signals)
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 β 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 $\beta_i(t_i, s_i) = a_i$.

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

$$
\begin{aligned}
&(1)\ \nu_j^{a_j,t_j}(t_{-j}, a_{-j}, \theta) > 0 \Rightarrow a_k \in BFR_k(t_k) \text{ for each } k \neq j; \\
&(2)\ \sum_{a_{-j}} \nu_j^{a_j,t_j}(t_{-j}, a_{-j}, \theta) > 0 \Rightarrow \pi_j(t_{-j}, \theta | t_j) > 0 \text{ for each } t_{-j}, \theta \text{; and} \\
&(3)\ a_j \in \arg\max_{a_j'} \sum_{t_{-j}, a_{-j}, \theta} \nu_j^{a_j,t_j}(t_{-j}, a_{-j}, \theta)\, u_j\left(\left(a_j', a_{-j}\right), \theta\right).
\end{aligned}
\tag{11}
$$

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

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

for some $\varepsilon > 0$. It is always possible to construct such an expansion for sufficiently small $\varepsilon > 0$ because of property (2) in (11) 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

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

For the converse, suppose that there exists an expansion $\left(S_j, \phi_j\right)_{j=1}^I$ of $\mathcal{T}$ and an equilibrium β 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 (10) 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 ϕ_i is independent of i. An expanded game $\left(G, \mathcal{T}, (S_i, \phi_i)_{i=1}^I\right)$ and a strategy profile β for that game will *induce* a decision rule $\sigma : T \times \Theta \to \Delta(A)$:

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

We record for completeness the corresponding results for expansions that satisfy the common prior assumption.

Proposition 9 (Informational Robustness to Common Prior Payoff-Irrelevant Signals)

If $\mathcal{T}$ is a common prior type space, then σ 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 β of $\left(G, \mathcal{T}, (S_i, \phi_i)_{i=1}^I\right)$ such that β induces σ.

A subjective version of Proposition 9 appears in Liu (2015) (and the common prior case is discussed in Section 5.2).

Proposition 10 (Informational Robustness to Common Prior Payoff-Relevant Signals)

If $\mathcal{T}$ is a common prior type space, then σ 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* β *of* $\left(G, \mathcal{T}, (S_i, \phi_i)_{i=1}^I\right)$ *such that* β *induces* σ.

Proposition 10 appears as Theorem 2 in Bergemann and Morris (2016a).

5 Discussion: Support Assumption and a Posteriori Equilibrium

In our informational robustness foundations, an expansion was characterized by each player's subjective belief about how all players' signals were being (stochastically) chosen as a function of players' types and the payoff state. Thus expansions were being explicitly identified with new signals that players observed. In this section, we will discuss an alternative way of describing an expansion of the type space, one that works directly with a player i's interim beliefs conditional on t_i and s_i. There are a number of reasons for doing so. First, this will highlight the significance and interpretation of the support assumption in the previous section. Second, it will clarify the connection to the prior literature. Finally, it will provide a step towards explaining the relation between "informational robustness" and "epistemic" foundations of solution concepts.

Suppose that we started with a type space $\mathcal{T} = (T_i, \pi_i)_{i=1}^I$ but now consider a different definition of an expanded type space (which will reduce to the previous one under additional assumptions). An expanded type space will take the form $\widetilde{\mathcal{T}} = \left(\widetilde{T}_i, \widetilde{\pi}_i\right)_{i=1}^I$ where $\widetilde{T}_i \subseteq T_i \times S_i$ and, for each i and $t_i \in T_i$, there exists $s_i \in S_i$ such that $\widetilde{t}_i = (t_i, s_i)$. What can we say about possible equilibrium behavior on such an expanded type space? We have built into this formulation the assumption that all possible types are rational with respect to some beliefs, and, in this sense, this formulation captures the idea of *a posteriori equilibrium*, the version of subjective correlated equilibrium introduced by Aumann (1974) and applied in Brandenburger and Dekel (1987). If we impose no restrictions on how the beliefs of (t_i, s_i) on the expanded type space relate to the beliefs of t_i on the original type space, then the original type space becomes irrelevant. In particular, say that an action is *ex post rationalizable* in the basic game $\mathcal{G}$ if it survives an iterative deletion procedure where, at each round, we delete actions that are not a best response given any conjecture over surviving actions and payoff states. Formally, let $EPR_i^0 = A_i$, let

EPR_i^{n+1} be the set of actions for which there exists $\nu_i \in \Delta(A_{-i} \times \Theta)$ such that

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

and let

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

This solution concept characterizes actions that can be played in equilibrium on any expanded type space if we dropped the support assumption from the analysis of the previous section.

This motivates putting additional restrictions on the expanded type space. We start by imposing a weak restriction that will correspond to the support assumption in the previous section: a player's beliefs on the original type space are not contradicted by his beliefs on the expanded type space. Thus

$$\sum_{s_{-i}} \widetilde{\pi}_i((t_{-i}, s_{-i}, \theta) | t_i, s_i) > 0 \Rightarrow \pi_i((t_{-i}, \theta) | t_i) > 0. \tag{12}$$

Restriction (12) reduces to the support restriction as defined in the previous section. Define

$$S_i(t_i) = \left\{ s_i \in S_i \,\middle|\, (t_i, s_i) \in \widetilde{T}_i \right\},$$

and

$$\phi_i((s_i, s_{-i}) \,|\, (t_i, t_{-i}), \theta) = \frac{1}{\# S_i(t_i)} \frac{\widetilde{\pi}_i((t_{-i}, s_{-i}, \theta) | t_i, s_i)}{\pi_i((t_{-i}, \theta) | t_i)} \tag{13}$$

whenever $\pi_i((t_{-i}, \theta) | t_i) > 0$ and $\phi_i(\cdot | (t_i, t_{-i}), \theta)$ is an arbitrary distribution otherwise. Now (12) implies

$$\sum_{s_{-i}} \phi_i((s_i, s_{-i}) \,|\, (t_i, t_{-i}), \theta)\, \pi_i((t_{-i}, \theta) | t_i) = \frac{1}{\# S_i(t_i)} \sum_{s_{-i}} \widetilde{\pi}_i((t_{-i}, s_{-i}, \theta) | t_i, s_i)$$

for each t_i and $s_i \in S_i(t_i)$, and so

$$\begin{aligned} \sum_{s_{-i}, t_{-i}, \theta} \phi_i((s_i, s_{-i}) \,|\, (t_i, t_{-i}), \theta)\, \pi_i((t_{-i}, \theta) | t_i) &= \frac{1}{\# S_i(t_i)} \sum_{s_{-i}, t_{-i}, \theta} \widetilde{\pi}_i((t_{-i}, s_{-i}, \theta) | t_i, s_i) \\ &= \frac{1}{\# S_i(t_i)} \\ &> 0, \end{aligned}$$

which is the support assumption. belief-invariance in the formulation of the previous section adds the requirement on the current expanded type space that

$$\sum_{s_{-i}} \widetilde{\pi}_i \left(t_{-i}, s_{-i}, \theta | t_i, s_i\right) = \widetilde{\pi}_i \left(t_{-i}, \theta | t_i\right)$$

for each i, t_{-i}, θ, t_i and s_i.

We noted earlier that a posteriori equilibrium from Aumann (1974) and Brandenburger and Dekel (1987) was equivalent to asking what can happen on all expanded type spaces in the case of complete information. But in the case of incomplete information – in the sense of there being many payoff states – we saw that the original type no longer mattered. Imposing either the weaker support assumption or the belief-invariance assumption are the natural generalizations of a posteriori equilibrium. Ex post rationalizability, like belief-free rationalizability and interim correlated rationalizability, reduces to correlated rationalizability in complete information games.

6 Appendix

Proof of Proposition 7. 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 $\nu_j^{a_j,t_j} \in \Delta(T_{-j} \times A_{-j} \times \Theta)$ such that

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

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

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

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 β 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 (10) 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 Advances in Theoretical Economics*, 1, Article 6.

Battigalli, P., and M. Siniscalchi (2003): "Rationalization and Incomplete Information," *BE Press 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.

——— (2009): "Robust Implementation in Direct Mechanisms," *Review of Economic Studies*, 76, 1175–1206.

——— (2012): *Robust Mechanism Design: The Role of Private Information and Higher Order Beliefs*. World Scientific Publishing.

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

——— (2015): "Informational Robustness and Solution Concepts," Discussion Paper 1973R, Cowles Foundation of Research, Yale University and Princeton University.

——— (2016a): "Bayes Correlated Equilibrium and the Comparison of Information Structures in Games," *Theoretical Economics*, 11, 487–522.

——— (2016b): "Information Design, Bayesian Persuasion, and Bayes Correlated Equilibrium," *American Economic Review Papers and Proceedings*, 106(5), 586–591.

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

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

Chen, J., S. Micali, and R. Pass (2015): "Tight Revenue Bounds with Possibilistic Beliefs and Level-k Rationality," *Econometrica*, 83(4), 1619–1639.

Dekel, E., D. Fudenberg, and D. Levine (2004): "Learning to Play Bayesian Games," *Games and Economic Behavior*, 46, 282–303.

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

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

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.

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

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

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. (2015): "Correlation and Common Priors in Games with Incomplete Information," *Journal of Economic Theory*, 157, 49–75.

MILGROM, P., AND N. STOKEY (1982): "Information, Trade and Common Knowledge," *Journal of Economic Theory*, 26, 17–27.

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

——— (1996): "Proximity of Information in Games with Incomplete Information," *Mathematics of Operations Research*, 21, 707–725.

MORRIS, S., AND C. SKIADAS (2000): "Rationalizable Trade," *Games and Economic Behavior*, 31, 311–323.

SEBENIUS, J., AND J. GEANAKOPLOS (1983): "Don't Bet on It: Contingent Agreements with Asymmetric Information," *Journal of the American Statistical Association*, 78, 424–426.

WEINSTEIN, J., AND M. YILDIZ (2007): "Impact of Higher-Order Uncertainty," *Games and Economic Behavior*, 60, 200–212.