Robustness of Bayesian Equilibria

Ronald Stauber*

March 16, 2004

Abstract

The standard model of a Bayesian game used in most applications assumes that players' beliefs are derived from a common knowledge prior on preference parameters. We analyze the robustness of equilibria of such games to perturbations in the information structure. In a type space environment (Harsanyi, 1967-68), we embed types corresponding to this information structure into an appropriately defined larger type space. We then perturb the embedded set using the notion of common p-belief (Monderer and Samet, 1989) by considering all types for which it is common p-belief that all players derive their beliefs about preference parameters from similar priors. For types in the perturbed set, we define an ε-equilibrium in which every player's strategy is an equilibrium strategy for the game where his individual prior is a common knowledge prior. Hence, this strategy is only a function of such a player's prior and private information, and does not depend on the exact form of his higher order beliefs. Based on this definition, we propose a notion of robustness that is independent of the specification of the underlying type space. This independence significantly simplifies the characterization of robust equilibria. The set of robust equilibria includes the set of ex post equilibria as a proper subset.

Keywords: Bayesian games, Robustness, Type spaces

JEL Classification: C72, D82

1 Introduction

A long-standing critique of the standard model of a Bayesian game is that beliefs about players' preferences are usually assumed to be derived from a common knowledge prior (CKP) on preference parameters, and are therefore common knowledge. One of the most prominent accounts of this critique is given by Wilson (1987), who

*Department of Economics, University of Illinois at Urbana-Champaign, 484 Wohlers Hall, 1206 S. Sixth Street, Champaign, Illinois 61820, E-mail: stauber@uiuc.edu

stresses that "game theory ... is deficient to the extent it assumes ... features to be common knowledge, such as one agent's probability assessment about another's preferences or information", and points out that "only by repeated weakening of common knowledge assumptions will the theory approximate reality". Although the model that assumes a common knowledge prior on preference parameters is less general than the original model of a Bayesian game introduced by Harsanyi (1967-68), it is used in most applications and textbook presentations. We will therefore refer to this model as the "textbook model", and to corresponding equilibria as "textbook equilibria".

It is sometimes argued that even though the assumption of a common knowledge prior almost never holds, the predictions of the textbook model would still be valid if the actual information structure is in some way "close" to the information structure of the textbook model. Our paper investigates this intuitive argument formally, by introducing a notion of robustness of textbook equilibria with respect to a precisely defined perturbation of the CKP assumption. To define robustness, we perturb beliefs while keeping equilibrium strategies fixed. One component of our definition of robustness is that the potential payoff gains from deviating from a robust equilibrium strategy are uniformly bounded by a small number whenever beliefs are "close" to CKP beliefs, and that this bound converges to zero as beliefs "approach" CKP beliefs. The intuition for this requirement is that a player who believes that the actual information structure is close to a CKP, does not make a big mistake as long as he plays according to a robust strategy. Hence, it may be rational for such a player to assume a CKP if the losses are small and if the cost of alternative modelling is large.

We use type spaces as a modelling framework, as suggested recently by Bergemann and Morris (2003) in the context of a robustness analysis of optimal mechanisms. The notion of a type space, introduced in the seminal work of Harsanyi (1967-68), provides an elegant solution to the problem of having to consider players' beliefs about other players' beliefs when analyzing games of incomplete information. Harsanyi's insight was to define a player's type to capture both his beliefs about all the specific characteristics of the game to be played, including all players' preferences, *and* his beliefs about other players' types. This formulation provides a mathematically tractable way to model beliefs about beliefs, and also allows for a very general definition of what a type can be. In fact, Mertens and Zamir (1985) showed that there exists a "universal type space", where types can be defined by any infinite hierarchy of beliefs about beliefs about preferences, as long as this hierarchy satisfies a simple consistency property.[1] Although the general definition of

[1] See Heifetz and Samet (1998) for a lucid review of the concept of a type space and an extension of Mertens and Zamir (1985).

a type space allows for great flexibility in defining a Bayesian game, for tractability reasons most of the literature in Economics employs the much narrower set of type spaces derived from a common knowledge prior on preference parameters. Because of this widespread reliance on textbook equilibria, it is important to determine to what extent the corresponding strategies are optimal if the CKP assumption is relaxed. This is the question that our notion of robustness attempts to answer.

The fact that a player's type captures all the characteristics of his beliefs in a type space model, suggests type spaces as the proper environment in which to relax the CKP assumption. To do this, we show how types corresponding to the textbook model can be embedded into an appropriately defined larger type space, in which a player's type is characterized by (1) a preference parameter which is private information, (2) an individual prior on all players' preference parameters, and (3) a probability measure characterizing his beliefs about the types of his opponents. We then define a perturbation of the embedded set by considering types whose beliefs are "close" to the beliefs of CKP types. To define such a closeness of belief structures, we use the notion common p-belief, which was introduced by Monderer and Samet (1989), and is generally recognized as an appropriate notion of approximate common knowledge. Thus, we consider those types whose beliefs are close to a CKP in the sense that they believe that their opponents' priors are close to their own prior with high probability, they believe with high probability that their opponents believe this with high probability, and so on ad infinitum. This approach allows us to examine the effects of relaxing the CKP assumption by looking at those types that belong to this perturbed set.[2] The construction of the perturbed set follows an idea of Kajii and Morris (1998), who show that a similar notion of approximate common knowledge is crucial for the lower hemicontinuity of the interim ε-equilibrium[3] correspondence. We discuss the relation between their paper and our work in the concluding section.

An advantage of using type spaces as a framework to define such a perturbation, is that we don't need to model explicitly the belief hierarchies involved in the definition. This is because in a type space, higher order beliefs are implicitly defined by the probability measures representing players' beliefs about their opponents' types. As a consequence, the perturbed set can be characterized by specifying necessary properties of such probability measures, without actually writing down the corresponding belief hierarchies.

[2]Note that we relax both the assumption that the players have a common prior on preference parameters and the assumption that the players' priors are common knowledge. However, this does not preclude the possibility that there is a common knowledge prior on the larger type space.

[3]A player's strategy is an ε-best response to the strategies of his opponents if the gains from deviating from this strategy are bounded by ε. A list of strategies such that each player's strategy is an ε-best response to his opponents' strategies, defines an ε-equilibrium.

We characterize robust equilibrium strategies by defining an ε-equilibrium for the type space game, such that each player with type/beliefs in the perturbed set described above plays an equilibrium strategy for the textbook game defined by his individual prior. If such an ε-equilibrium exists, the corresponding equilibrium strategies will satisfy the requirements for robustness mentioned above. To define the ε-equilibrium, we first introduce a topology on the set of all pairs of priors on preference parameters and corresponding textbook equilibria. Intuitively, this topology captures strategic closeness between textbook equilibria corresponding to various priors, in the sense that two equilibria are close if there is a bound on the gains from deviating from the strategies prescribed by one of the equilibria, as long as a player's opponents play according to the other equilibrium. Since these payoff incentives to deviate decrease with the magnitude of this bound, we can construct a neighborhood base for this "strategic topology" by varying the bound. We then introduce an "equilibrium map" as a map from priors to corresponding textbook equilibria that is uniformly continuous according to this topology, and show that such maps define the ε-equilibrium.

A player's strategy as specified by such an ε-equilibrium is an equilibrium strategy for the textbook equilibrium corresponding to his individual prior, as determined by the equilibrium map. Such a strategy only depends on a player's prior and his private information preference parameter, and is independent of the exact functional form of the player's higher order beliefs. The only restrictions on the players' higher order beliefs we make is that it is common p-belief that all players' priors are not too different, i.e., that the players' types belong to the perturbed set defined above. Thus, as long as this assumption is satisfied, these strategies are independent of variations in second and higher order beliefs. Intuitively, such a property is desirable, since it is conceivable that players may not be very confident about their beliefs about other players' beliefs, the beliefs about other players' beliefs about their own beliefs, and so on. This invariance to changes in higher order beliefs introduces an element of max-min decision theory into our definition of equilibrium, since the prescribed strategies constitute an ε-best response for all higher order beliefs corresponding to some type in the perturbed set.

We believe that this invariance to small changes in higher order beliefs, together with the bound on the gains from deviating, are both essential for a characterization of robustness. Since both these properties are satisfied if an equilibrium lies on an equilibrium map, we can define a robust textbook equilibrium as an equilibrium that lies on some equilibrium map.

The motivation for this approach is that players may be confident that the common knowledge assumption is approximately satisfied, but may otherwise be uncertain about the exact distribution of their opponents' beliefs. For such players,

it may not be worth considering complicated type spaces when calculating their strategies, if the gains from doing this are bounded by a small number.

An additional benefit of this formulation is that the independence of the prescribed strategies from higher order beliefs greatly simplifies the identification of robust equilibria. Since equilibrium maps are only defined in terms of priors and corresponding textbook equilibria, they can be characterized without use of the specific type space postulated in the analysis.

Given any textbook equilibrium corresponding to some prior, there need not exist an equilibrium map such that the given equilibrium lies on this map. In this case, there may exist types in the perturbed set defined above, for which the gains from deviating from a strategy prescribed by this equilibrium would yield gains exceeding ε. Whether this is the case would depend on the exact form of this type's higher order beliefs. Nonetheless, such strategies cannot possess the robustness properties described above. Thus, not all textbook equilibria are robust, and therefore our notion of robustness yields a proper refinement for textbook equilibria.

After showing how robust equilibria can be defined in relation to an equilibrium map for a general model of a Bayesian game, we consider the case of finite games. For such games, we derive a precise characterization of equilibrium maps in terms of the Euclidean topology on strategies. Specifically, this characterization implies that equilibrium maps can be discontinuous and do not require lower hemicontinuity of the equilibrium correspondence for existence.

The paper is structured as follows: Section 2 presents an example and uses it to illustrate some problems that arise if the CKP assumption is relaxed. Section 3 reviews some background material on Bayesian games, type spaces and common p-belief. Section 4 derives our characterization of robustness. Properties of robust equilibria are collected in section 5. Section 6 presents additional examples illustrating various properties of equilibrium maps. Section 7 relates our approach to the literature and discusses possible extensions and applications.

2 An Example

The following example helps illustrate the main ideas of the paper and some problems that can arise if the common knowledge assumption is relaxed. The analysis of the general model can be followed without reading this section, if one is willing to ignore some references to the example.

Example 1: The example is taken from Engl (1995), who attributes it to Rubinstein (1989). We modify Engl's version by considering a simpler information

structure. Because the textbook equilibria based on common knowledge of beliefs are a main ingredient of our analysis, we start by characterizing all such equilibria.

Two players, denoted by 1 and 2, are playing one of two coordination games, A or B. The set of available actions, $\{a, b\}$, is the same for both players and in both games. The payoffs depend on the players' choices and the game that is played, as shown in Figure 1.

		Player 2 a	Player 2 b
Player 1	a	$1, 1$	$0, -1$
	b	$-1, 0$	$0, 0$

A

		Player 2 a	Player 2 b
Player 1	a	$0, 0$	$0, -1$
	b	$-1, 0$	$1, 1$

B

Figure 1: Player 2 believes that game A is played with probability π.

Only player 1 knows the true game; Player 2 believes that game A is played with probability π and game B with probability $1-\pi$. The assumption that beliefs are common knowledge implies that 2 knows that 1 knows the true game, that 1 knows the value of π, that both players know they know, and so on.

In order to allow for mixed strategy equilibria, we characterize the players' strategies using the probabilities with which they choose action a. For player 2, we denote this probability by σ_2. Since player 1 knows the true game, he can condition his action on the actual game, so we let $\sigma_1(A)$ be the probability that 1 chooses action a if the true game is A, and $\sigma_1(B)$ the probability that 1 chooses a if the true game is B.

The (Bayesian) Nash equilibria for this game are as follows:

1. $\sigma_2 = \sigma_1(A) = \sigma_1(B) = 1$, for $\pi \in [0, 1]$;
2. $\sigma_2 = 0$, $\sigma_1(A) \in [0, 1]$, $\sigma_1(B) = 0$, for $\pi \in [0, \frac{1}{3}]$;
3. $\sigma_2 = 0$, $\sigma_1(A) \in \left[0, \frac{1-\pi}{2\pi}\right]$, $\sigma_1(B) = 0$, for $\pi \in [\frac{1}{3}, 1]$;
4. $\sigma_2 = \frac{1}{2}$, $\sigma_1(A) = 1$, $\sigma_1(B) = \frac{1}{2} - \frac{\pi}{1-\pi}$, for $\pi \in [0, \frac{1}{3}]$;
5. $\sigma_2 \in [0, \frac{1}{2}]$, $\sigma_1(A) = 1$, $\sigma_1(B) = 0$, for $\pi = \frac{1}{3}$.[4]

[4]The first equilibrium is straightforward: If player 2 chooses action a, player 1's best response in both games is to also choose a. If 1 chooses a in both games, 2's best response is to also choose a.

The second and third equilibria follow from the observation that if 2 chooses b, then 1 will choose b in game B, and will be indifferent between a and b if the true game is A. If 1 plays action a with probability $\sigma_1(A)$, then 2 will choose b only if $\sigma_1(A) \leq \frac{1-\pi}{2\pi}$. Note that this also implies that

Since the set of equilibria varies as a function of π, we can look at the equilibrium correspondence $\Gamma : [0,1] \to 2^{[0,1]^3}$, $\pi \mapsto \Gamma(\pi)$, which maps π to the set of all equilibria $\sigma := (\sigma_1(A), \sigma_1(B), \sigma_2)$ of the game in which it is common knowledge that player 1 knows the true game and player 2 believes that game A is played with probability π. It is well-known that the equilibrium correspondence need not be lower hemicontinuous. This is illustrated in the example at $\pi = 1/3$ by considering the equilibria of type 5 above, which belong to the set $\Gamma' := \{\sigma \in [0,1]^3 \mid \sigma_1(A) = 1,\ \sigma_1(B) = 0,\ \text{and } \sigma_2 \in (0, \frac{1}{2}]\}$.[5] We use these observations later, in our discussion of robustness issues.

Relaxing the common knowledge assumption: To illustrate the difficulties that arise if the common knowledge assumption does not hold, we maintain the assumption that 1 knows the true game, but relax the assumption that 2's beliefs are common knowledge. One way to motivate this is to assume that before 1 receives his private information, both players have some historical data on the basis of which they estimate the probability that game A will be played. If the data or estimation procedures are not identical, the players' estimates will most likely differ. Denote 1's estimate by π_1 and 2's estimate by π_2. Given appropriate assumptions on the differences in data and estimation procedures, π_1 could also be interpreted as 1's assessment of 2's beliefs π_2.

If the difference between π_1 and π_2 is small, would a rational player who is aware of this, but does not know the exact beliefs of his opponent, ever consider playing a strategy as prescribed by one of the equilibria derived above, calculated by setting π equal to his own estimate π_i? Specifically, if we denote such an equilibrium by σ^{π_i}, would the strategy pairs $(\sigma_i^{\pi_i}, \sigma_{-i}^{\pi_{-i}})$ constitute an ε-equilibrium for the game where players are aware of differences in beliefs? More precisely, to determine whether σ^{π_i} is robust, we ask whether given any $\varepsilon > 0$, there exists a $\delta > 0$, such that whenever $|\pi_i - \pi_{-i}| < \delta$, player i's gains from changing his strategy relative

for all $\pi \in [0,1]$ there exists a pure strategy equilibrium with $\sigma_2 = \sigma_1(A) = \sigma_1(B) = 0$.
In order to get the fourth equilibrium, assume 2 randomizes between the two actions, i.e. $\sigma_2 > 0$. If the true game is A, player 1 then picks action a. If the true game is B, player 1 picks a if $\sigma_2 > 1/2$, b if $\sigma_2 < 1/2$, and is indifferent between a and b if $\sigma_2 = 1/2$. Noting that player 2 is indifferent between the two actions only if $\sigma_1(B) = \frac{1}{2} - \frac{\pi}{1-\pi}$ gives the fourth equilibrium.
Note also that only for $\pi = 1/3$ does there exist an equilibrium where player 2 randomizes with $\sigma_2 \neq 1/2$, for if $\sigma_2 > 1/2$, player 1 would pick a in both games, in which case player 2 would also choose a with probability 1; if instead $\sigma_2 < 1/2$, player 1 picks a in game A and b in game B, which implies that 2's best response is to choose a if $\pi \geq 1/3$ and b if $\pi \leq 1/3$. Thus, only if $\pi = 1/3$, is 2 indifferent between the two actions, whence we get the last equilibrium.

[5] To see this, note that for any sequence $\pi_n \searrow 1/3$, there do not exist equilibria $\sigma_n \in \Gamma(\pi_n)$ converging to an element of Γ'. If $\pi = \frac{1}{3} + \varepsilon$ and $\sigma_2 > 1/2$, player 1 would choose a in game B, so $\sigma_1(B) = 1$, whereas if $0 < \sigma_2 \leq 1/2$, we still have $\sigma_1(A) = 1$ and $\sigma_1(B) = 0$, but then player 2's best response would be to choose a (i.e. $\sigma_2 = 1$), a contradiction.

to $\sigma_i^{\pi_i}$ are bounded by ε.

To support such an ε-equilibrium, any player i would thus need to know that, at least with a high probability, π_{-i} is such that $|\pi_i - \pi_{-i}| < \delta$, and that $\sigma_i^{\pi_i}$ is an ε-best response to $\sigma_{-i}^{\pi_{-i}}$. Therefore, in order for this player to play according to the strategy $\sigma_i^{\pi_i}$, the players must coordinate on a selection from the equilibrium correspondence, i.e., a function mapping π to an equilibrium σ^π, with the property that $(\sigma_i^{\pi_i}, \sigma_{-i}^{\pi_{-i}})$ constitutes an ε-equilibrium whenever $|\pi_i - \pi_{-i}| < \delta$. This observation motivates the definition of an "equilibrium map" as such a selection. We want this map to define an ε-equilibrium for the game when beliefs are close to common knowledge. However, for this to be the case, it is necessary not only that both players believe that $|\pi_i - \pi_{-i}| < \delta$ with high probability, but also that they know that their opponents know this, that their opponents know they know and so on.

It thus seems that we either have to analyze an infinite hierarchy of beliefs or impose common knowledge of second- or higher-order beliefs. Fortunately, the notion of common p-belief allows us to only assume an approximate form of common knowledge, which will be sufficient to deal with this problem. In addition, by applying common p-belief in the framework of a type space, we do not have to worry about infinite belief hierarchies. We will formalize this approach in the context of a general model.

Robustness issues: For now, suppose that an equilibrium map as defined above does yield an ε-equilibrium for the common knowledge game. If this is the case, a textbook equilibrium will be robust if it lies on some equilibrium map. What would such a map look like in the example? The easiest choice would be a constant map where players always choose action a or action b independent of their assessment of π. Another example would be one where the probabilities $(\sigma_1(A), \sigma_1(B), \sigma_2)$ vary continuously with π.

The key question becomes, for any given π, which equilibria can be included in such a map? Since equilibria for which this is not the case do not satisfy our robustness requirements, the answer to this question yields an equilibrium refinement for the textbook model. In the example, an examination of the set of equilibria shows that only the equilibria in the set Γ' at which we have shown that lower hemicontinuity of the equilibrium correspondence fails, *cannot* lie on any continuous selection from the equilibrium correspondence. Hence, these equilibria are the only candidates for equilibria that fail to be robust.

Interestingly, this not the case: All equilibria $\gamma \in \Gamma'$ are robust in the sense that for all such γ there exists an equilibrium map with $\sigma^{1/3} = \gamma$. To see this, let

$\gamma_2 \in (0, \frac{1}{2}]$, and define an equilibrium map σ^π as follows:

$$\sigma^\pi = \begin{cases} (1, \frac{1}{2} - \frac{\pi}{1-\pi}, \frac{1}{2}), & \text{if } \pi < \frac{1}{3} \\ (1, 0, \gamma_2), & \text{if } \pi = \frac{1}{3} \\ (\frac{1-\pi}{2\pi}, 0, 0), & \text{if } \pi > \frac{1}{3} \end{cases}. \tag{1}$$

Note that there is a discontinuity in σ_2, but that $\sigma_1(A)$ and $\sigma_1(B)$ are continuous as a function of π. Since at $\pi_1 = \frac{1}{3}$ player 1's best response is the same for all $\sigma_2 \in [0, \frac{1}{2}]$, he has no incentive to deviate from the strategy prescribed by σ^π. For $\pi_2 = \frac{1}{3}$ and $\pi_1 = \frac{1}{3} \pm \delta$, player 2's losses from playing according to σ^{π_2} converge to zero as $\delta \to 0$, as long as player 1 follows σ^{π_1}. This implies that the map defined above has the required properties of an equilibrium map.

Thus, in this example, all equilibria are robust. We will see later this is not the case for all games.

3 Preliminaries

3.1 A Model of a Bayesian Game

Bayesian Games are commonly defined as a list $\{\mathcal{I}, (V_i)_{i\in\mathcal{I}}, (A_i)_{i\in\mathcal{I}}, (u_i)_{i\in\mathcal{I}}\}$, where:

- $\mathcal{I}$ is a finite set of players [$i \in \mathcal{I}$, $-i := \mathcal{I} \setminus \{i\}$]; and for all i
- V_i is a measurable set of player i's payoff types [$v_i \in V_i$, $V := \prod_{i\in\mathcal{I}} V_i$, $v \in V$];
- A_i is a set of available actions for i [$a_i \in A_i$, $A := \prod_{i\in\mathcal{I}} A_i$, $a \in A$]; and
- $u_i : V \times A \to \mathbb{R}$ is i's utility function, such that $(v, a) \mapsto u_i(v, a)$.

In order to simplify the exposition, we only consider two-player games, $\mathcal{I} = \{1, 2\}$. Nevertheless, the ideas and methods of the paper generalize easily to the case of more players.

The usual way to proceed in defining a Bayesian game is to assume that each player i knows his own payoff type and has beliefs about other players' payoff types given by a probability distribution over V_{-i}. In most textbook expositions and applications, these beliefs are derived from a prior distribution on V, which is assumed to be both a common prior and common knowledge. If this is the case, a player's payoff type uniquely determines his beliefs. For example, if $\pi \in \Delta V$ is the common knowledge prior, then player i's beliefs are given by the conditional of π given v_i. Thus the "type" of a player in the textbook model, what we called a player's preference parameter in the introduction, corresponds to a player's payoff type in the current model.

Harsanyi's main insight was that without assuming a common knowledge distribution on V, we can define a player's type to include his private information about both preferences and beliefs. In our model, players' types will have both preference and belief components. We therefore distinguish between a player's payoff type, which as a parameter may affect all players' preferences or utilities, and his type, which may include additional information.

We allow for such types by augmenting our model of a Bayesian game with the following definition of a type space, which is a variant of the definitions given in Heifetz and Samet (1998) and Bergemann and Morris (2003):

Definition 1 *A* `type space` *is a triple* $\langle T, \nu, \mu \rangle \equiv \langle (T_i)_{i \in \mathcal{I}}, (\nu_i)_{i \in \mathcal{I}}, (\mu_i)_{i \in \mathcal{I}} \rangle$*, such that for each player* i

1. T_i *is a measurable space with associated* σ*-algebra* $\mathscr{T}_i$*;*
2. ν_i *is a measurable function* $\nu_i : T_i \to V_i$*;*
3. μ_i *is a measurable function* $\mu_i : T_i \to \Delta T$*, where* ΔT *is equipped with the* σ*-algebra generated by all sets of the form* $\beta^p(E) = \{\mu' \in \Delta T \,|\, \mu'(E) \geq p\}$*, with* $E \in \mathscr{F} := \prod_{i \in \mathcal{I}} \mathscr{T}_i$ *and* $p \in [0, 1]$*;*
4. *The marginal of* $\mu_i(t_i)$ *on* T_i *is* δ_{t_i}*, the measure in* ΔT_i *concentrated at* t_i*.*

T_i is the set of all possible types of player i. Given a player's type t_i, we interpret $\nu_i(t_i)$ as his payoff type v_i, and $\mu_i(t_i)$ as his beliefs about the distribution of all players' types. Property 4 states that each player knows his own type.[6] It also implies that $\mu_i(t_i)$ defines a probability measure on T_{-i}, denoted by $\mu_i^c(t_i)$, by letting $\mu_i^c(t_i)(E_{-i}) = \mu_i(t_i)(T_i \times E_{-i})$ for all $E_{-i} \in \mathscr{T}_{-i}$.

Note that we do not assume that beliefs are derived from a common prior. This would be the case if there exists a distribution function $\mu \in \Delta T$ such that the conditional of μ given t_i is equal to $\mu_i^c(t_i)$.

By including a type space, we get a complete characterization of a Bayesian game as a list $\{\mathcal{I}, (V_i)_{i \in \mathcal{I}}, (A_i)_{i \in \mathcal{I}}, (u_i)_{i \in \mathcal{I}}, \langle T, \nu, \mu \rangle\}$, which is assumed to be common knowledge. This formulation includes the model in which beliefs are derived from a common knowledge prior $\pi \in \Delta V$ as a special case, with $T_i = V_i$, ν_i the identity map, and μ_i the conditional of π given v_i.[7]

[6] Alternatively, we could define $\mu_i(t_i)$ as a measure on T_{-i}, but the current definition is more convenient when working with common p-belief.

[7] For other examples of type spaces, including type spaces based on infinite hierarchies of beliefs, see Bergemann and Morris (2003).

3.2 Common p-Belief in a Type Space Environment

For an event to be common knowledge, it is necessary that both players know that the event is true, that both player know that their opponents know this, that both players know that their opponents know they know, and so on. The intuition behind common p-belief is that even though the players may not know something with complete certainty, they might still believe that it is true with high probability. Hence, Monderer and Samet (1989) replace the infinite hierarchy of statements about knowledge with an infinite hierarchy of statements about belief with sufficiently high probability p. Thus, common p-belief of some event means that both players believe it to be true with probability p, that both players believe with probability p that their opponents believe that it is true with probability p and so on.

In the context of our problem, we use common p-belief to relax the assumption that beliefs about payoff types are common knowledge, by assuming instead that it is common p-belief that players' beliefs about the distribution of payoff types are not too different. In Example 1, this would mean that it is common p-belief that the difference between π_1 and π_2 does not exceed some small number δ.

Since a player's type in a type space determines his beliefs about the distribution of his opponent's type, it implicitly determines his beliefs about the distribution of his opponent's beliefs. A type space therefore provides a natural framework in which to define common p-belief. This section confirms that the definition and characterization of common p-belief, which were originally formulated for a state space model with a common prior, also hold for a type space without a common prior. The proofs of the stated results are straightforward extensions of proofs in Monderer and Samet (1989) and Kajii and Morris (1997), and can be found in the appendix.

A p-belief operator for player i is a map $B_i^p : \mathscr{F} \to 2^T$, such that for all $E \in \mathscr{F}$:

$$B_i^p(E) := \{(t_i, t_{-i}) \in T \,|\, \mu_i(t_i)(E) \geq p\}.$$

We interpret $B_i^p(E)$ as the set of all type combinations for which player i believes that the event E will occur with probability p or higher.

As noted by Heifetz and Samet (1998), $B_i^p(E) = \mu_i^{-1}(\beta^p(E)) \times T_{-i}$, which implies that $B_i^p(E) \in \mathscr{F}$ for any $E \in \mathscr{F}$. Denote the σ-algebra on T generated by μ_i^{-1} by $\mathscr{F}_i$: An element of $\mathscr{F}_i$ is therefore a set of the form $\mu_i^{-1}(\beta^p(E)) \times T_{-i}$ for some $E \in \mathscr{F}$, and so $B_i^p(E) \in \mathscr{F}_i$.

The following proposition collects various properties of B_i^p:

Proposition 1 *For every $E, F \in \mathscr{F}$,*

(i) $B_i^p(E) \in \mathscr{F}_i$;

(ii) $E \in \mathscr{F}_i \Rightarrow E = B_i^p(E)$;

(iii) $E \subseteq F \Rightarrow B_i^p(E) \subseteq B_i^p(F)$;

(iv) *If* $\{E_n\}_n \subseteq \mathscr{F}$ *is a decreasing sequence of events such that* $E_{n+1} \subseteq E_n$, *then* $B_i^p(\bigcap_n E_n) = \bigcap_n B_i^p(E_n)$;

(v) $E \in \mathscr{F}_i \Rightarrow B_i^p(E \cap F) = E \cap B_i^p(F)$.

Now define $B^p(E) := \bigcap_{i \in \mathcal{I}} B_i^p(E)$. $B^p(E)$ is the set of types $t \in T$ for which all players believe that E occurs with probability p or higher. Note that since $\mathcal{I}$ is countable, $B^p(E) \in \mathscr{F}$. We can apply B^p iteratively any number of times n to get new sets $[B^p]^n(E) \in \mathscr{F}$. Let $\mathcal{C}^p(E) := \bigcap_{n \geq 1} [B^p]^n(E)$, so $\mathcal{C}^p(E) \in \mathscr{F}$. This is the set of all types such that all players believe that E occurs with probability of at least p, all players believe that all other players believe that E occurs with probability of at least p and so on.

In addition to defining p-belief operators and common p-beliefs, one of the main contributions of Monderer and Samet (1989) was to show that the iterative definition of common p-belief as formulated by the operator $\mathcal{C}^p$ is equivalent to a fixed point definition using `evident` p `-belief events`, which are those events $E \in \mathscr{F}$ for which $E \subseteq B^p(E)$. Analogously to the well known fixed point definition of common knowledge, evident p-belief events can be used to define common p-belief:

Definition 2 *An event* $E \in \mathscr{F}$ *is* `common` p `-belief at` $t \in T$ *if there exists an evident* p*-belief event* F *such that* $t \in F$ *and* $F \subseteq B^p(E)$.

The following results imply that $\mathcal{C}^p(E)$ is exactly the set of types for which E is common p-belief:

Lemma 1 *For* $E \in \mathscr{F}$, $B^p(B^p(E)) \subseteq B^p(E)$.

Lemma 2 *For* $E \in \mathscr{F}$, $\mathcal{C}^p(E) \subseteq B^p(\mathcal{C}^p(E))$, *i.e.,* $\mathcal{C}^p(E)$ *is an evident* p*-belief event.*

Proposition 2 $E \in \mathscr{F}$ *is common* p*-belief at* $t \Leftrightarrow t \in \mathcal{C}^p(E)$.

The operator $\mathcal{C}^p$ allows us to define a perturbation of the common knowledge prior assumption in a precise way, and thus plays a central role in our characterization of robustness.

4 A Characterization of Robustness

4.1 Separable Type Spaces

In the previous section we gave a standard definition of a type space. To facilitate the comparison with the textbook model, we now introduce some additional assumptions on the type spaces that will be considered in the subsequent analysis.

We start by augmenting our definition of a Bayesian game with a set of possible states of the world Θ, and assume that the realized state $\theta \in \Theta$ determines the true distribution over V according to which the players' payoff types are drawn. Hence, for each θ, there exists a corresponding distribution $F(\theta) \in \Delta V$, so Θ represents a parametrization of the distributions over V that are regarded as feasible by the players, or respectively, by the modeler. For example, the set $\{F(\theta)\}_{\theta \in \Theta}$ could be chosen to only include distributions over V where the individual v_i's are independently distributed. We let Θ be a measurable set, but do not make any additional assumptions at this stage.

The players are not assumed to know the true state in Θ, so in order to make a decision, they must form beliefs about the distribution of the states, beliefs about their opponent's beliefs and so on. In such an environment, the common knowledge prior assumption of the textbook model would imply that, without taking into account the private information about payoff types, it is common knowledge that all players have identical beliefs about the distribution of the θ's. Such an assumption is obviously quite strong. Harsanyi (1967-68) provides a defense of the common prior assumption in the context of a general type space. He argues that the prior should only reflect information that is public and hence common to all players, and that all differences in the players' assessments of the distribution of states should be regarded as resulting from differences in players' interpretation of the public information or any additional private information they might have. Therefore, all such differences should be reflected in the players' types, and all players should share a common prior over such types. But for this argument to be valid, a player's type must include not only information about his payoff type, but also any *private* information that this player might have about the state of nature, *and* any objective or subjective beliefs he might have about the other player's type. Hence, the common prior assumption could be regarded as reasonable for a general type space, but less so for a type space that only includes payoff types, as in the textbook model.

In extending the textbook model, we maintain the implicit assumption that the only information the players have about the state of nature θ is public information. We model this by assuming that there exists a public signal $s \in S$ that is observed by all players and is correlated with θ. Players are not assumed to have any information about θ except for this signal and the private knowledge of their own

payoff type v_i. The difference from the textbook model is that we allow players to use different models or inference processes to interpret the public signal s, which implies that different players may have different assessments about the distribution of θ's, and therefore about the distribution of payoff types. We do not require the individual models to be common knowledge, so each player must form beliefs about the model used by his opponent, about the opponent's beliefs about the player's model, and so on.

We model this interconnected belief structure by introducing, for each player i, a measurable set B_i of possible `belief types` $b_i \in B_i$. Analogously to the type space introduced in the previous section, a type b_i captures both information about the model used by player i to interpret the signal s, and his beliefs about the belief types of his opponents. We thus assume that there exist measurable functions $\pi_i : B_i \times S \to \Delta V$ and $\varphi_i : B_i \to \Delta B$, with the property that the marginal of $\varphi_i(b_i)$ on B_i is δ_{b_i}. Hence, $\varphi_i(b_i)$ is a probability measure on B concentrated at b_i. We use $\varphi_i^c(b_i)$ to denote the probability measure on B_{-i} induced by $\varphi_i(b_i)$. We interpret $\pi_i(b_i, s)$ as type b_i's assessment of the distribution of payoff types for the whole population, when he observes signal s. Note that while we do assume that $\varphi_i(b_i)$ is such that each player knows his own belief type, we do not assume that the marginal of $\pi_i(b_i, s)$ is concentrated at v_i. Thus, $\pi_i(b_i, s)$ is this player's prior distribution over payoff types, before learning his own payoff type. We introduce these measures as the counterpart of the common knowledge priors of the textbook model. Just as with any prior on V, $\pi_i(b_i, s)$ defines a player's beliefs about his opponent's payoff type as the conditional of $\pi_i(b_i, s)$ given v_i, denoted by $\pi_i(b_i, s)(\cdot|v_i)$. Note also that the functions $(\varphi_i)_{i\in\{1,2\}}$ allow us to define common p-belief for subsets of $B = B_1 \times B_2$.

A player's private information now consists of his belief type b_i and his payoff type v_i. We therefore introduce the notation $t_i := (b_i, v_i)$ for player i's type, and let $T_i := B_i \times V_i$. Together with the observed public signal s, each player's type t_i determines his beliefs about his opponent's type $t_{-i} = (b_{-i}, v_{-i})$, given by the product measure $\varphi_i^c(b_i) \times \pi_i(b_i, s)(\cdot|v_i)$ on $B_{-i} \times V_{-i}$. An important implication of this, is that from the perspective of each type t_i of player i, the belief types and payoff types of his opponent are independently distributed. This independence stems from our assumption that players' belief types contain no information about the state of nature that determines the distribution of payoff types. Independence is not crucial for our results, but it significantly simplifies the model and notation. We will mention later how it could be relaxed without affecting our definition of robustness.

Taking a parametrization $\{F(\theta)\}_{\theta\in\Theta}$ of the feasible distributions over payoff types as given, we can define a `separable type space` as a list $\langle S, B, V, \varphi, \pi\rangle$.

Note that in terms of the notation of the previous section, we have $\nu_i(t_i) = v_i$ and $\mu_i^c(t_i) = \varphi_i^c(b_i) \times \pi_i(b_i, s)(\cdot|v_i)$.

From now on, we only consider separable type spaces. We also assume that the sets V_i, B_i and A_i are separable metric spaces which admit a complete metric, and that utility functions are bounded, in the sense that there exists a positive number M such that $|u_i(v, a) - u_i(v', a')| \leq M$, for all $i \in \mathcal{I}$, $v, v' \in V$ and $a, a' \in A$.

4.2 Definitions

We study interim equilibria in behavioral strategies, where interim means that the players' objective is to maximize their utilities conditional on their types. A `behavioral strategy` for player i is defined as a map $\sigma_i : \mathbf{A_i} \times T_i \times S \to [0, 1]$, where $\mathbf{A_i}$ denotes the Borel σ-field of A_i, $\sigma_i(\cdot|t_i, s) : \mathbf{A_i} \to [0, 1]$ is a probability measure for all $(t_i, s) \in T_i \times S$, and $\sigma_i(D|\cdot) : T_i \times S \to [0, 1]$ is measurable for every $D \in \mathbf{A_i}$.[8]

We let

$$U_i(v_i, v_{-i}, \sigma_i(t_i, s), \sigma_{-i}(t_{-i}, s)) := \int_A u_i(v_i, v_{-i}, a_i, a_{-i})\sigma_i(da_i|t_i, s)\sigma_{-i}(da_{-i}|t_{-i}, s), \tag{2}$$

and

$$W_i(s, v_i, \sigma_i, \sigma_{-i}) := \int_{B_{-i}} \int_{V_{-i}} U_i(v_i, v_{-i}, \sigma_i, \sigma_{-i}(b_{-i}, v_{-i}, s))\pi_i(b_i, s)(dv_{-i}|v_i)\varphi_i^c(b_i)(db_{-i}). \tag{3}$$

Hence, U_i denotes player i's expected utility given both players' types and strategies, and W_i denotes the expected value of U_i given i's beliefs.

If the game is defined using some common knowledge prior $\pi \in \Delta V$, we define $U_i(v_i, v_{-i}, \sigma_i(v_i), \sigma_{-i}(v_{-i}))$ as in (2), and

$$W_i(\pi, v_i, \sigma_i, \sigma_{-i}) := \int_{V_{-i}} U_i(v_i, v_{-i}, \sigma_i(v_i), \sigma_{-i}(v_{-i}))\pi(dv_{-i}|v_i). \tag{4}$$

We also use $W_i(\pi, v_i, a_i, a_{-i})$ when the strategies are such that a_i and a_{-i} are played with probability 1.

A `Bayesian Nash equilibrium` for the type space game is a strategy pair (σ_1^*, σ_2^*), such that for all i, s, t_i and σ_i',

$$W_i(s, v_i, \sigma_i^*, \sigma_{-i}^*) \geq W_i(s, v_i, \sigma_i', \sigma_{-i}^*).$$

To get the definition of an ε-`equilibrium`, the inequality is replaced by

$$W_i(s, v_i, \sigma_i^*, \sigma_{-i}^*) \geq W_i(s, v_i, \sigma_i', \sigma_{-i}^*) - \varepsilon.$$

[8] See Milgrom and Weber (1985) for additional details on the definition of behavioral strategies and the relation between behavioral, mixed and distributional strategies in Bayesian games.

For the textbook game defined by π, we denote a Bayesian Nash equilibrium by $(\sigma_1^\pi, \sigma_2^\pi)$, so $W_i(\pi, v_i, \sigma_i^\pi, \sigma_{-i}^\pi) \geq W_i(\pi, v_i, \sigma_i', \sigma_{-i}^\pi)$ for all i, v_i and σ_i'.

4.3 Types with Approximate Common Knowledge

Just as beliefs about payoff types are derived from the common prior in the textbook model, they are derived from $\pi_i(b_i, s)$ for any belief type b_i in a separable type space. We can thus embed types corresponding to a common knowledge game into a separable type space using the functions $(\pi_i)_{i \in \mathcal{I}}$. Since a player's belief type b_i determines both his assessment of the distribution of payoff types, $\pi_i(b_i, s)$, and his beliefs about his opponent's belief type b_{-i}, this embedding will only depend on his belief type and not on his payoff type v_i. To be precise, we do not worry about probability zero events, so the common knowledge types we consider are actually all types for which common knowledge holds with probability 1.

An obvious requirement for such common knowledge types is that they believe with probability 1 that their opponent's belief type is such that $\pi_{-i}(b_{-i}, s) = \pi_i(b_i, s)$. Since the values of the π_i's are probability measures, we must specify what it means for two measures to be equal. Which notion of equality is appropriate will depend on the specification of the game. For example, if the sets V_i are finite, we could use the Euclidean topology on the conditional probabilities of the π's given the players' payoff types. If the utility functions are continuous in payoff types, we could measure equality using the weak topology on the players' conditional distributions.[9]

To avoid making any additional assumptions at this stage, we introduce a topology that seems appropriate for the general model considered so far. For any $\pi \in \Delta V$, we define a neighborhood base at π for this topology by the sets:

$$\mathcal{N}(\pi, \gamma) = \left\{ \pi' \in \Delta V \,\middle|\, \sup_{i, v_i, a_i, \sigma_{-i}} \left| W_i(\pi, v_i, a_i, \sigma_{-i}) - W_i(\pi', v_i, a_i, \sigma_{-i}) \right| < \gamma \right\},$$

where $\gamma > 0$. Because this topology is defined in terms of the utility functions for the game that is being analyzed, it yields convergence of expected utilities in the sense that a sequence π_n converges to π in this topology only if

$$d^e(\pi_n, \pi) := \sup_{i, v_i, a_i, \sigma_{-i}} |W_i(\pi_n, v_i, a_i, \sigma_{-i}) - W_i(\pi, v_i, a_i, \sigma_{-i})|$$

converges to zero. It is easy to see that d^e defines a metric on the set ΔV. Although we use this metric for the remainder of this section, it is important to note that none of the results depend on the specification of d^e, and that any alternative notion of equality could have been used instead.

[9] See Chapter 3 of Stroock (1999) for an introduction to various topologies for sets of measures.

Now define

$$\mathcal{A}^0 := \left\{(b_i, b_{-i}) \in B \,\middle|\, d^e(\pi_i(b_i, s), \pi_{-i}(b_{-i}, s)) = 0,\ \forall s\right\}.$$

Thus, $\mathcal{A}^0$ is the set of all belief type pairs that have the same assessments of the distribution of payoff types, independent of the observed public signal s. Consider the belief operator B_i^1 applied to the set $\mathcal{A}^0$:

$$B_i^1(\mathcal{A}^0) = \left\{(b_i, b_{-i}) \in B \,|\, \varphi_i(b_i)(\mathcal{A}^0) = 1\right\}.$$

Note that the belief operator is now applied to subsets of B and not T. Since the marginal of $\varphi_i(b_i)$ is concentrated at b_i, a type $b_i \in \text{proj}_{B_i} B_i^1(\mathcal{A}^0)$ puts probability 1 on types b_{-i} with $d^e(\pi_i(b_i, s), \pi_{-i}(b_{-i}, s)) = 0$, $\forall s$. Recall that $B^1(\mathcal{A}^0) = \bigcap_{i\in\mathcal{I}} B_i^1(\mathcal{A}^0)$. Then the set $\mathcal{C}^1(\mathcal{A}^0) = \bigcap_{n\geq 1}[B^1]^n(\mathcal{A}^0) \subset B$ of types for which $\mathcal{A}^0$ is common 1-belief, is the set of those belief types that believe with probability 1 that their opponents assessments is the same as their own, they believe with probability 1 that their opponents believe this with probability 1 and so on. Therefore, defining $\mathcal{P}^0 := \mathcal{C}^1(\mathcal{A}^0)$, we get those belief type pairs for which it is common knowledge that both players have the same assessments, i.e., the set of belief types corresponding to some common knowledge prior game.

The results of Heifetz and Samet (1998) on the existence of a universal type space imply that we can always find a type space for which this embedding of common knowledge prior types is not empty.

The definition of the perturbation of $\mathcal{P}^0$ parallels the definition of $\mathcal{P}^0$. Our objective is to get a subset of B, the set of belief type pairs, which contains those types for which the common knowledge prior assumption holds approximately.

Define, for any small $\delta > 0$, the set of belief type pairs for which the difference in assessments is bounded by δ:

$$\mathcal{A}^\delta := \left\{(b_i, b_{-i}) \in B \,\middle|\, d^e(\pi_i(b_i, s), \pi_{-i}(b_{-i}, s)) \leq \delta,\ \forall s\right\}.$$

Instead of considering belief types for which $\mathcal{A}^\delta$ is common knowledge, or equivalently, common 1 -belief, we include in the perturbation those belief types for which $\mathcal{A}^\delta$ is common $(1-\delta)$-belief. Thus, we apply the belief operator $B_i^{1-\delta}$ to the set $\mathcal{A}^\delta$ to get

$$B_i^{1-\delta}(\mathcal{A}^\delta) = \left\{(b_i, b_{-i}) \in B \,|\, \varphi_i(b_i)(\mathcal{A}^\delta) \geq 1-\delta\right\}.$$

As before, we have $B^{1-\delta}(\mathcal{A}^\delta) = \bigcap_{i\in\mathcal{I}} B_i^{1-\delta}(\mathcal{A}^\delta)$ and $\mathcal{C}^{1-\delta}(\mathcal{A}^\delta) = \bigcap_{n\geq 1}[B^{1-\delta}]^n(\mathcal{A}^\delta)$. To get the perturbation of $\mathcal{P}^0$, we just define $\mathcal{P}^\delta := \mathcal{C}^{1-\delta}(\mathcal{A}^\delta)$.

Let $\mathcal{P}_i^\delta := \text{proj}_{B_i}\mathcal{P}^\delta$. We use the following properties of $\mathcal{P}^\delta$ in the proof of our main result. These properties follow directly from the definition as a common $(1-\delta)$ -belief event:

(i) $\forall b_i \in \mathcal{P}_i^\delta$, $\varphi_i(b_i)(\mathcal{A}^\delta) \geq 1-\delta$;

(ii) $\forall b_i \in \mathcal{P}_i^\delta$, $\varphi_i(b_i)(\mathcal{P}^\delta) \geq 1-\delta$;

(iii) $\forall b_i \in \mathcal{P}_i^\delta$, $\varphi_i(b_i)(\mathcal{P}^\delta \cap \mathcal{A}^\delta) \geq 1-2\delta$.

Thus, if $b_i \in \mathcal{P}_i^\delta$, no matter what signal s player i observes, he believes that $d^e(\pi_i(b_i,s), \pi_{-i}(b_{-i},s)) \leq \delta$ with high probability, he believes that his opponent also believes this to be true with high probability, and so on.

4.4 Equilibria with Approximate Common Knowledge

By definition, for each $b_i \in \mathcal{P}_i^0$, an equilibrium strategy for the textbook game defined by the prior $\pi_i(b_i,s)$ is exactly optimal. In the game defined by a separable type space, we now identify an ε-equilibrium for players with belief type $b_i \in \mathcal{P}_i^\delta$, such that each player's strategy is an equilibrium strategy for the textbook game defined by the prior $\pi_i(b_i,s)$. Note that any such prior must belong to the convex hull of the set of distributions in ΔV defined by the set of states Θ. Denote the convex hull by $\text{co}_\Theta\{F(\theta)\}$.

Since payoffs are determined by both players' strategies, and their assessments π_i and π_{-i} are allowed to differ in our model, it is necessary to define an equilibrium as a map $\pi \mapsto \sigma^{[\pi]}$, which maps a probability measure $\pi \in \text{co}_\Theta\{F(\theta)\}$ to an equilibrium of the textbook game defined by π. As seen in Example 1, not every map will work. To characterize the requirements that we need to impose on such maps, we define, for a given Bayesian game, a topology on the set

$$\mathcal{E} := \{\sigma^\pi \,|\, \pi \in \text{co}_\Theta\{F(\theta)\} \wedge \sigma^\pi \text{ is an equilibrium of the textbook game defined by } \pi\}.$$

Intuitively, this topology measures strategic closeness between elements of $\mathcal{E}$, so we refer to it as the `strategic topology`. We define this strategic topology by specifying a neighborhood base for any $\sigma^\pi \in \mathcal{E}$ given by the following family of sets:

$$\mathcal{S}(\sigma^\pi, \xi) := \{\sigma^{\bar{\pi}} \in \mathcal{E} \,\big|\, \sup_{i, v_i, \sigma_i'} [W_i(\pi, v_i, \sigma_i', \sigma_{-i}^{\bar{\pi}}) - W_i(\pi, v_i, \sigma_i^\pi, \sigma_{-i}^{\bar{\pi}})] < \xi\},$$

where $\xi > 0$ and $\sigma_i' \in \Delta A_i$. Thus, $\sigma^{\bar{\pi}} \in \mathcal{S}(\sigma^\pi, \xi)$ if all players' expected gains from switching to a strategy different from σ_i^π are bounded by ξ, given that their opponents play strategy $\sigma_{-i}^{\bar{\pi}}$ and the expected utilities are calculated using π.

The following definition of an equilibrium map is crucial in our characterization of robustness:

Definition 3 *A map $\pi \mapsto \sigma^{[\pi]}$ is an* `equilibrium map` *if for all $\xi > 0$, there exists a $\delta > 0$, such that $\sigma^{[\bar{\pi}]} \in \mathcal{S}(\sigma^{[\pi]}, \xi)$ whenever $d^e(\pi, \bar{\pi}) < \delta$.*

To provide a better understanding of this definition, we use the idea behind the construction of the sets $\mathcal{S}(\sigma^\pi, \xi)$ to define a distance function on the set $\mathcal{E}$. For any σ^π and $\sigma^{\bar{\pi}}$ in $\mathcal{E}$, let

$$\vec{d}(\sigma^\pi, \sigma^{\bar{\pi}}) := \sup_{i, v_i, \sigma_i'} [W_i(\pi, v_i, \sigma_i', \sigma_{-i}^{\bar{\pi}}) - W_i(\pi, v_i, \sigma_i^\pi, \sigma_{-i}^{\bar{\pi}})].$$

As explained above, $\vec{d}(\sigma^\pi, \sigma^{\bar{\pi}})$ measures how close $\sigma^{\bar{\pi}}$ is to σ^π, but not vice versa. This is because $\vec{d}(\sigma^\pi, \sigma^{\bar{\pi}})$ is not necessarily a symmetric function.[10] We therefore define

$$d^s(\sigma^\pi, \sigma^{\bar{\pi}}) := \max\{\vec{d}(\sigma^\pi, \sigma^{\bar{\pi}}), \vec{d}(\sigma^{\bar{\pi}}, \sigma^\pi)\}$$

as a measure of strategic closeness between σ^π and $\sigma^{\bar{\pi}}$. Note that d^s does not constitute a metric on $\mathcal{E}$ because it does not satisfy the triangle inequality.[11] In addition, the topology defined by d^s is not a Hausdorff topology and is therefore not metrizable.[12] The following lemma shows that the definition of an equilibrium map can be rephrased using the function d^s:

Lemma 3 *A map $\pi \mapsto \sigma^{[\pi]}$ is an equilibrium map if and only if for all $\xi > 0$, there exists a $\delta > 0$, such that $d^s(\sigma^{[\pi]}, \sigma^{[\bar{\pi}]}) < \xi$ whenever $d^e(\pi, \bar{\pi}) < \delta$.*

Thus, an equilibrium map is a map which is uniformly continuous using the metric d^e on $\text{co}_\Theta\{F(\theta)\}$ and the distance d^s on $\mathcal{E}$.

The following proposition shows that this kind of uniform continuity is sufficient to define strategies that constitute an ε-best response for all belief types in $\mathcal{P}^\delta$:

Proposition 3 *Given any equilibrium map $\pi \mapsto \sigma^{[\pi]}$ and $\xi > 0$, let $\delta > 0$ be determined by Definition 3. For all i, let $\hat{\sigma}_i$ be a behavioral strategy such that $\hat{\sigma}_i(\cdot|b_i, v_i, s) = \sigma_i^{[\pi_i(b_i,s)]}(\cdot|v_i)$ for $b_i \in \mathcal{P}_i^\delta$, and $\hat{\sigma}_i(\cdot|b_i, v_i, s)$ equal to some arbitrary strategy for $b_i \notin \mathcal{P}_i^\delta$. Then for all players with belief type $b_i \in \mathcal{P}_i^\delta$, all s, and all $\sigma_i' \in \Delta A_i$, $W_i(s, v_i, \hat{\sigma}_i, \hat{\sigma}_{-i}) > W_i(s, v_i, \sigma_i', \hat{\sigma}_{-i}) - (\xi + 4\delta M)$.*

Proof: The idea of the proof is straightforward: We first bound the losses of a player with $b_i \in \mathcal{P}_i^\delta$ from the fact that an opponent with $b_{-i} \notin \mathcal{P}_{-i}^\delta$ plays an

[10] To see this, consider Example 1 and look at the equilibrium strategies $(1,0,0)^{\frac{1}{3}}$ and $(1,1,1)^{\frac{1}{3}}$, where the superscripts indicate that $\pi = \frac{1}{3}$. Then $\vec{d}((1,0,0)^{\frac{1}{3}}, (1,1,1)^{\frac{1}{3}}) = \frac{4}{3}$ and $\vec{d}((1,1,1)^{\frac{1}{3}}, (1,0,0)^{\frac{1}{3}}) = 1$.

[11] This is again easily established by looking at Example 1 and the equilibria $\sigma^I = (1,0,0)^{\frac{1}{3}}$, $\sigma^{II} = (1,0,\frac{1}{2})^{\frac{1}{3}}$ and $\sigma^{III} = (1,1,1)^{\frac{1}{3}}$. We have $d^s(\sigma^I, \sigma^{II}) = 0$, $d^s(\sigma^{II}, \sigma^{III}) = \frac{2}{3}$ and $d^s(\sigma^I, \sigma^{III}) = \frac{4}{3}$.

[12] The constant sequence given by σ^I, as defined in the previous footnote, converges both to σ^I and σ^{II}.

unspecified strategy. We then bound the losses from the fact that an opponent with $b_{-i} \in \mathcal{P}^\delta_{-i}$ may have an assessment $\pi_{-i}(b_{-i}, s)$ which is different from $\pi_i(b_i, s)$. Combining the two bounds completes the proof.

For any $E \subseteq B$, denote the complement of E by $\overline{E}$, i.e., $\overline{E} := T \setminus E$.

For all $b_i \in \mathcal{P}^\delta_i$ and all σ_i,

$$W_i(s, v_i, \sigma_i, \hat{\sigma}_{-i}) = \int_{B_{-i}} \int_{V_{-i}} U_i(v_i, v_{-i}, \sigma_i, \hat{\sigma}_{-i}) \pi_i(b_i, s)(dv_{-i}|v_i) \varphi^c_i(b_i)(db_{-i})$$

$$\begin{aligned} = \quad & \varphi^c_i(b_i)(\mathcal{P}^\delta \cap \mathcal{A}^\delta) \int_{B_{-i}} \left\{ \int_{V_{-i}} U_i(v, \sigma_i, \hat{\sigma}_{-i}) \pi_i(b_i, s)(dv_{-i}|v_i) \right\} \varphi^c_i(b_i)(db_{-i}|\mathcal{P}^\delta \cap \mathcal{A}^\delta) + \\ & + \varphi^c_i(b_i)(\overline{\mathcal{P}^\delta \cap \mathcal{A}^\delta}) \int_{B_{-i}} \left\{ \int_{V_{-i}} U_i(v, \sigma_i, \hat{\sigma}_{-i}) \pi_i(b_i, s)(dv_{-i}|v_i) \right\} \varphi^c_i(b_i)(db_{-i}|\overline{\mathcal{P}^\delta \cap \mathcal{A}^\delta}) = \\ = \quad & \int_{B_{-i}} \int_{V_{-i}} U_i(v, \sigma_i, \hat{\sigma}_{-i}) \pi_i(b_i, s)(dv_{-i}|v_i) \varphi^c_i(b_i)(db_{-i}|\mathcal{P}^\delta \cap \mathcal{A}^\delta) + \\ & + \varphi^c_i(b_i)(\overline{\mathcal{P}^\delta \cap \mathcal{A}^\delta}) \left[\int_{B_{-i}} \left\{ \int_{V_{-i}} U_i(v, \sigma_i, \hat{\sigma}_{-i}) \pi_i(b_i, s)(dv_{-i}|v_i) \right\} \varphi^c_i(b_i)(db_{-i}|\overline{\mathcal{P}^\delta \cap \mathcal{A}^\delta}) - \right. \\ & \left. - \int_{B_{-i}} \left\{ \int_{V_{-i}} U_i(v, \sigma_i, \hat{\sigma}_{-i}) \pi_i(b_i, s)(dv_{-i}|v_i) \right\} \varphi^c_i(b_i)(db_{-i}|\mathcal{P}^\delta \cap \mathcal{A}^\delta) \right]. \end{aligned}$$

It follows that

$$\left| W_i(s, v_i, \sigma_i, \hat{\sigma}_{-i}) - \int_{B_{-i}} \int_{V_{-i}} U_i(v, \sigma_i, \hat{\sigma}_{-i}) \pi_i(b_i, s)(dv_{-i}|v_i) \varphi^c_i(b_i)(db_{-i}|\mathcal{P}^\delta \cap \mathcal{A}^\delta) \right| \leq 2\delta M, \tag{5}$$

where we have used the bound on utilities and property (iii) of $\mathcal{P}^\delta$.

For $b_{-i} \in \mathcal{P}^\delta_{-i}$, we have defined $\hat{\sigma}_{-i} = \sigma^{[\pi_{-i}(b_{-i}, s)]}_{-i}$, which implies that we can write

$$\int_{V_{-i}} U_i(v, \sigma_i, \hat{\sigma}_{-i}) \pi_i(b_i, s)(dv_{-i}|v_i) = W_i(\pi_i(b_i, s), v_i, \sigma_i, \sigma^{[\pi_{-i}(b_{-i}, s)]}_{-i}).$$

Since conditional on $\mathcal{P}^\delta \cap \mathcal{A}^\delta$, we have $d^e(\pi_i(b_i, s), \pi_{-i}(b_{-i}, s)) < \delta$, the fact that $\sigma^{[\pi]}$ is an equilibrium map yields

$$W_i(\pi_i(b_i, s), v_i, \sigma^{[\pi_i(b_i, s)]}_i, \sigma^{[\pi_{-i}(b_{-i}, s)]}_{-i}) > W_i(\pi_i(b_i, s), v_i, \sigma'_i, \sigma^{[\pi_{-i}(b_{-i}, s)]}_{-i}) - \xi,$$

for all σ'_i. Taking expectations on both sides of the previous equation using $\varphi^c_i(b_i)(db_{-i}|\mathcal{P}^\delta \cap \mathcal{A}^\delta)$, we get

$$\begin{aligned} & \int_{B_{-i}} \int_{V_{-i}} U_i(v, \hat{\sigma}_i, \hat{\sigma}_{-i}) \pi_i(b_i, s)(dv_{-i}|v_i) \varphi^c_i(b_i)(db_{-i}|\mathcal{P}^\delta \cap \mathcal{A}^\delta) > \\ & > \int_{B_{-i}} \int_{V_{-i}} U_i(v, \sigma'_i, \hat{\sigma}_{-i}) \pi_i(b_i, s)(dv_{-i}|v_i) \varphi^c_i(b_i)(db_{-i}|\mathcal{P}^\delta \cap \mathcal{A}^\delta) - \xi, \end{aligned} \tag{6}$$

for all $b_i \in \mathcal{P}_i^\delta$ and all σ_i'.

Finally, applying (5), (6), and (5) again, we have

$$\begin{aligned}
&W_i(s, v_i, \hat{\sigma}_i, \hat{\sigma}_{-i}) \geq \\
&\geq \int_{B_{-i}} \int_{V_{-i}} U_i(v, \hat{\sigma}_i, \hat{\sigma}_{-i}) \pi_i(b_i, s)(dv_{-i}|v_i) \varphi_i^c(b_i)(db_{-i}|\mathcal{P}^\delta \cap \mathcal{A}^\delta) - 2\delta M > \\
&> \int_{B_{-i}} \int_{V_{-i}} U_i(v, \sigma_i', \hat{\sigma}_{-i}) \pi_i(b_i, s)(dv_{-i}|v_i) \varphi_i^c(b_i)(db_{-i}|\mathcal{P}^\delta \cap \mathcal{A}^\delta) - \xi - 2\delta M \geq \\
&\geq W_i(s, v_i, \sigma_i', \hat{\sigma}_{-i}) - \xi - 4\delta M,
\end{aligned}$$

for all $b_i \in \mathcal{P}_i^\delta$ and all σ_i'. ◀

Proposition 3 does not specify a strategy for belief types $b_i \notin \mathcal{P}_i^\delta$. We could have instead required such types to choose a best response given their beliefs, or some strategy constituting an ε-best response. Since our goal is to analyze the case where the CKP is approximately satisfied, this is not relevant for the subsequent analysis. Moreover, calculating a best response for beliefs defined by infinite hierarchies is a non-trivial task, so a result that does not depend on players having to solve such a problem seems preferable.

4.5 Defining Robustness

For belief types $b_i \in \mathcal{P}_i^\delta$, Proposition 3 states that the strategies prescribed by the equilibrium map are an ε-best response, with $\varepsilon = \xi + 4\delta M$. Thus, for these types, the maximum potential loss from choosing such a strategy is bounded by ε, independently of the actual prior $\pi_i(b_i, s)$. In addition, the strategies only depend on the priors, i.e., the players' first order beliefs, and their payoff types v_i. Therefore, a player who believes that $b_i \in \mathcal{P}_i^\delta$, i.e., who thinks that the CKP assumption is approximately satisfied, does not need to worry about the details of his higher order beliefs – the gains from deviating from the prescribed strategy cannot exceed ε for any infinite belief hierarchy such that $b_i \in \mathcal{P}_i^\delta$. All this is a consequence of the definition of an equilibrium map, which requires that

$$W_i(\pi_i(b_i, s), v_i, \sigma_i^{[\pi_i(b_i,s)]}, \sigma_{-i}^{[\pi_{-i}(b_{-i},s)]}) > W_i(\pi_i(b_i, s), v_i, \sigma_i', \sigma_{-i}^{[\pi_{-i}(b_{-i},s)]}) - \xi,$$

for all σ_i' and all b_{-i} such that $d^e(\pi_i(b_i, s), \pi_{-i}(b_{-i}, s)) < \delta$. Examining the proof of Proposition 3 shows that this requirement is stronger than necessary. For the proof to work, it would be enough for the expectation of this inequality, using the measure $\varphi_i^c(b_i)(db_{-i}|\mathcal{P}^\delta \cap \mathcal{A}^\delta)$, to hold. By instead using the stronger formulation, we implicitly require that the expectation holds for all distributions $\varphi_i^c(b_i)(db_{-i}|\mathcal{P}^\delta \cap \mathcal{A}^\delta)$. This strengthens the conclusion of Proposition 3, since we don't need any

assumption on the distributions $\varphi_i(b_i)$ except for $b_i \in \mathcal{P}_i^\delta$. Hence, the optimality of the equilibrium strategies is robust to a wide range of higher order beliefs that a player may have. Another important advantage of this formulation is that it allows us to define equilibrium maps without reference to the type space used to analyze the game. This independence from type space considerations greatly simplifies the characterization of equilibrium maps and robust equilibria.

An alternative interpretation of this result is as follows: A player only knows that it is common $(1-\delta)$-belief that his opponent's assessment of the distribution of payoff types satisfies $d^e(\pi_i(b_i,s),\pi_{-i}(b_{-i},s)) < \delta$, but either has no additional information about the exact distribution of his opponent's belief type b_{-i}, or is not very confident in his own beliefs. The strategy prescribed by an equilibrium map guarantees that if such a player considers the worst-case opponent belief $\pi_{-i}(b_{-i},s)$, he will not lose more than ε relative to his optimal strategy. This is reminiscent of a max-min decision rule, where players choose an action that maximizes the worst-case outcome as a function of his opponent's assessment $\pi_{-i}(b_{-i},s)$.

The actual choice of a δ would depend on a player's degree of confidence in the CKP assumption. It follows from the definition of an equilibrium map and Proposition 3 that the maximum gains from deviating converge to zero as $\delta \to 0$, i.e., that $\varepsilon \to 0$ as $\delta \to 0$. Thus, a player only needs to know that δ is close to zero – this continuity of the bound ε around CKP beliefs insures that he will not make any big mistake by choosing the strategy prescribed by the equilibrium map.

The previous discussion shows that if a textbook equilibrium lies on some equilibrium map, then the corresponding strategies satisfy the two robustness requirements discussed in the introduction. This motivates the following definition:

Definition 4 *An equilibrium of a textbook game is a* `robust equilibrium` *if it lies on some equilibrium map.*

We will show later that many textbook equilibria are robust, but that there also exist non-robust equilibria. Thus, robustness is a non-trivial refinement.

Defining robustness in terms of equilibrium maps has the advantage that in order to determine whether a textbook equilibrium is robust, we only need to look at the equilibrium correspondence that maps priors to corresponding textbook equilibria. If the equilibrium lies on a selection from this correspondence that is an equilibrium map, the equilibrium is robust. If such a selection does not exist, it is not. The alternative of analyzing the equilibrium correspondence of the larger type space game is obviously a much more arduous task.

Two properties of equilibrium maps need further motivation – the continuity (using d^s) at all points of the domain $\text{co}_\Theta\{F(\theta)\}$, and specifically, the uniform continuity. The necessity of continuity follows from the fact that we want the

prescribed strategies to be an ε-best response for all higher order beliefs that satisfy $b_i \in \mathcal{P}_i^\delta$. For an example of what can happen if continuity is violated, consider Example 1 and the map

$$\tilde{\sigma}^\pi = \begin{cases} (1,1,1), & \text{if } \pi \leq \frac{1}{2} \\ (0,0,0), & \text{if } \pi > \frac{1}{2} \end{cases}.$$

$\tilde{\sigma}^\pi$ is continuous at all $\pi \neq \frac{1}{2}$, but if for example, player i always believes that his opponent's beliefs π_{-i} are distributed uniformly on $[\pi_i, \pi_i + \delta]$, then for all $\delta > 0$ there is always a positive (Lebesgue) measure of π_i's that will deviate from the prescribed strategy as long as ξ is small enough. Our limited assumptions on higher order beliefs then imply that all players with $\pi_i \leq \frac{1}{2}$ may have an incentive to deviate.

The uniform continuity is important because it allows the conditions on the required degree of common knowledge, measured in terms of δ, to be stated independently of a player's prior π_i. Without the uniformity, it would by possible that the δ required for some ε-equilibrium converges to zero when π_i approaches some π^*. But this would imply that only common knowledge of π^* can support a textbook equilibrium for this prior.

As mentioned after the introduction of separable type spaces, it is possible to relax the assumption of independence between belief types and payoff types, without affecting our definition of equilibrium. This can be done by assuming instead, that even though belief types may be correlated with the state of nature that governs the distribution of payoff types, the effect on priors of such private information is outweighed by the public information contained in the signal s. Specifically, we can replace the independence assumption with a bound on the difference between a player's prior $\pi_i(b_i, s)$ if he only knows his own belief type b_i, and his prior conditioned on any opponent's belief type b_{-i} such that $(b_i, b_{-i}) \in \mathcal{A}^\delta$. As long as the bound is less than the ξ given in the statement of Proposition 3, the original proof only needs an additional step to take care of this bound, in order to carry over to the relaxed case.

A stronger notion of robust equilibria which has been analyzed extensively in the recent literature in game theory and mechanism design, is that of an ex post equilibrium.[13] An `ex post equilibrium` is a strategy pair that represents a Nash equilibrium for all possible realization of the vector of payoff types. Thus, a strategy prescribed by an ex post equilibrium is optimal for all beliefs an agent might have about the distribution of payoff types, which implies that

[13] See for example, Bergemann and Morris (2003) and Chung and Ely (2002).

Proposition 4 *Any ex post equilibrium σ^{xp} defines an equilibrium map by setting $\sigma^{[\pi]} = \sigma^{\text{xp}}$ for all π.*

An ex post equilibrium is therefore robust according to our definition. In fact, the equilibrium map for Example 1 defined in equation (1) shows that there exist equilibrium maps that are not defined by an ex post equilibrium, so the set of ex post equilibria is a strict subset of the set of robust equilibria.

5 Properties of Equilibrium Maps

This section characterizes equilibrium maps. For this purpose, we assume that the sets of payoff types and actions, V_i and A_i, are finite, and that $\text{co}_\Theta\{F(\theta)\} \equiv \Delta V$. Although the results are derived for the finite case, we expect them to generalize to games in which V_i and A_i are compact metric spaces.

The finiteness assumption and the fact that equilibrium maps are defined in terms of textbook equilibria as π varies over ΔV, allow us to define a behavioral strategy for player i as a map $\sigma_i : V_i \to \Delta^{|A_i|}$, where $\Delta^{|A_i|} \subset \mathbb{R}^{|A_i|}$ denotes the $(|A_i| - 1)$-dimensional simplex. As before, we use σ^π to denote an equilibrium strategy for the game defined by a common knowledge prior π, and $\sigma_i^\pi(v_i)$ to denote the corresponding strategy of a player i of type v_i.

For two vectors $x, y \in \mathbb{R}^l$, we use the `sum metric` $|x - y| = \sum_{j=1}^{l} |x_j - y_j|$ to analyze convergence in the Euclidean topology. If $\pi \in \Delta V$, we denote the conditional of π given v_i by $\pi[v_i]$, so $\pi[v_i] \in \Delta^{|V_{-i}|}$. Equality of two measures π and π' in ΔV can now be defined in terms of the convergence of the conditionals $\pi[v_i]$. We thus substitute the metric d^e in the definition of an equilibrium map with the following metric:

$$d(\pi, \pi') := \max_{i, v_i} |\pi[v_i] - \pi'[v_i]|. \tag{7}$$

It is easy to see that convergence in d implies convergence in d^e for all utility functions.

Since the topology defined by the strategic distance d^s is not Hausdorff, we start the characterization of equilibrium maps by relating the strategic closeness of equilibria to the Euclidean convergence of strategies as measured by the metric

$$d(\sigma, \sigma') := \max_{i, v_i} |\sigma_i(v_i) - \sigma_i'(v_i)|.$$

We use the same d as in (7) since both metrics are based on the sum metric in Euclidean space.

For any map $\pi \mapsto \sigma^{[\pi]}$ we can now consider two notions of continuity, one using strategic closeness of strategies as defined by d^s, and the other using closeness in

the Euclidean topology as defined by the metric d. From now on we reserve the term `continuity` for the latter notion and refer to the uniform continuity required in the definition of an equilibrium map as `strategic continuity`. Similarly, by convergence of a strategy profile we mean convergence in the metric d.

The following proposition shows that continuity of a map $\pi \mapsto \sigma^{[\pi]}$ implies strategic continuity. The converse does not hold, as seen from Example 1. Thus, continuity is sufficient, but not necessary, for a map $\pi \mapsto \sigma^{[\pi]}$ to be an equilibrium map.

Proposition 5 *Any continuous map $\pi \mapsto \sigma^{[\pi]}$ is an equilibrium map.*

Proof: First, note that since ΔV is compact, the continuity of $\sigma^{[\pi]}$ implies uniform continuity. Thus, for each $\epsilon > 0$ there exists a $\delta > 0$ such that $d(\pi, \pi') < \delta \Rightarrow d(\sigma^{[\pi]}, \sigma^{[\pi']}) < \epsilon$. Denote the δ corresponding to any ϵ by $\delta(\epsilon)$. We can assume w.l.o.g. that $\delta(\epsilon)$ is non-decreasing in ϵ and that $\delta(\epsilon) \leq \epsilon$ for all $\epsilon > 0$.

Since $\sigma^{[\pi']}$ is an equilibrium for the game defined by the common knowledge prior π', for each i, v_i and $\tilde{\sigma}_i \in \Delta^{|A_i|}$,

$$W_i(\pi', v_i, \sigma_i^{[\pi']}, \sigma_{-i}^{[\pi']}) \geq W_i(\pi', v_i, \tilde{\sigma}_i, \sigma_{-i}^{[\pi']}).$$

Now given some $\epsilon > 0$, $d(\sigma^{[\pi]}, \sigma^{[\pi']}) < \epsilon$ implies that $|\sigma_i^{[\pi]}(v_i) - \sigma_i^{[\pi']}(v_i)| < \epsilon$, so

$$W_i(\pi', v_i, \sigma_i^{[\pi]}, \sigma_{-i}^{[\pi']}) > W_i(\pi', v_i, \sigma_i^{[\pi']}, \sigma_{-i}^{[\pi']}) - \epsilon M > W_i(\pi', v_i, \tilde{\sigma}_i, \sigma_{-i}^{[\pi']}) - \epsilon M.$$

Similarly, $d(\pi, \pi') < \delta(\epsilon)$ implies that

$$W_i(\pi, v_i, \sigma_i^{[\pi]}, \sigma_{-i}^{[\pi']}) > W_i(\pi, v_i, \tilde{\sigma}_i, \sigma_{-i}^{[\pi']}) - (\epsilon + 2\delta(\epsilon))M.$$

Strategic continuity follows by letting $\xi \equiv (\epsilon + 2\delta(\epsilon))M$ for any $\xi > 0$. ◀

Recalling that the equilibrium correspondence is upper hemicontinuous (Milgrom and Weber, 1985) gives the following corollary:

Corollary 1 *If there exists a unique equilibrium for all $\pi \in \Delta V$, then the equilibrium correspondence is an equilibrium map.*

Although the equilibrium correspondence need not be lower hemicontinuous, a well-known result about correspondences states that any upper hemicontinuous correspondence with domain and range equal to some complete metric space is continuous on a residual.[14] A `residual` of a metric space is a countable intersection of dense open subsets. Baire's Category Theorem states that a residual is dense.[15]

[14] A formal statement and proof of this result can be found in Aubin and Frankowska (1990).

[15] This theorem is proved in most advanced real analysis textbooks, e.g. Folland (1999).

In our model of a Bayesian game, this implies that the equilibrium correspondence is continuous on a residual. It follows from the definition of continuity that the set of continuity points of the equilibrium correspondence is a dense, open subset of ΔV. This observation suggests a possible strategy for finding equilibrium maps: find a piecewise continuous function that is strategically continuous at all discontinuity points. The following proposition characterizes such points.

Proposition 6 *Let $BR_i^{\pi}(\cdot)$ denote player i's best response correspondence in the game defined by the common knowledge prior π, and suppose that $\pi \mapsto \sigma^{[\pi]}$ is piecewise continuous. Then $\pi \mapsto \sigma^{[\pi]}$ is an equilibrium map iff for all discontinuity points π^*, all i, and all sequences π^n and π^m converging to π^* such that the limits $\lim_{\pi^n \to \pi^*} \sigma_i^{[\pi^n]}$ and $\lim_{\pi^m \to \pi^*} \sigma_i^{[\pi^m]}$ exist, we have*

1. $\sigma_i^{[\pi^*]} \in BR_i^{\pi^*}(\lim_{\pi^n \to \pi^*} \sigma_{-i}^{[\pi^n]})$,
2. $\lim_{\pi^n \to \pi^*} \sigma_i^{[\pi^n]} \in BR_i^{\pi^*}(\sigma_{-i}^{[\pi^*]})$, *and*
3. $\lim_{\pi^n \to \pi^*} \sigma_i^{[\pi^n]} \in BR_i^{\pi^*}(\lim_{\pi^m \to \pi^*} \sigma_{-i}^{[\pi^m]})$.

Since Proposition 6 is a direct corollary of the next proposition, we do not give a separate proof. Properties 1 and 2 are actually a consequence of property 3. We only include them in the statement of the proposition since they emphasize specific properties of discontinuous equilibrium maps. Note also that the upper hemicontinuity of the equilibrium correspondence implies that $\lim_{\pi^n \to \pi^*} \sigma^{[\pi^n]}$ is an equilibrium for the game defined by the common knowledge prior π^*.

The idea behind Proposition 6 can be used to derive a necessary and sufficient condition for a map to be an equilibrium map. This condition can be formulated in terms of the Euclidean topology on strategies, without reference to the strategic topology. In order to state this result, we first introduce additional definitions and notation:

Given a map $\pi \mapsto \sigma^{[\pi]}$, its `graph` $G(\sigma)$ is the subset of $\Delta V \times \prod_i \Delta^{|A_i|}$ defined by

$$G(\sigma) := \left\{ (\pi, \sigma') \in \Delta V \times \prod_i \Delta^{|A_i|} \;\middle|\; \sigma' = \sigma^{[\pi]} \right\}.$$

The closure of this set, $\overline{G(\sigma)}$, defines a correspondence $\Sigma : \Delta V \to \prod_i \Delta^{|A_i|}$ by letting $G(\Sigma) \equiv \overline{G(\sigma)}$. By definition, $\sigma' \in \Sigma(\pi)$ if and only if there exists a sequence $\pi^n \to \pi$ such that $\sigma^{[\pi^n]} \to \sigma'$. As noted above, if Γ denotes the equilibrium correspondence, then $\sigma' \in \Sigma(\pi)$ implies that $\sigma' \in \Gamma(\pi)$. Since $\prod_i \Delta^{|A_i|}$ is compact and $G(\Sigma)$ is closed, Σ is upper hemicontinuous.

For any set $Y \subset \prod_i \Delta^{|A_i|}$, the `upper inverse` of Y by the correspondence Σ, $\Sigma^+[Y]$, is defined by

$$\Sigma^+[Y] := \{\pi \in \Delta V \mid \Sigma(\pi) \subset Y\}.$$

We will use the fact that a compact-valued correspondence is upper hemicontinuous if and only if the upper inverse of any open set is also open.[16]

We can now state our characterization of equilibrium maps:

Proposition 7 *Given a map $\pi \mapsto \sigma^{[\pi]}$, let Σ be defined by $G(\Sigma) \equiv \overline{G(\sigma)}$. Then $\sigma^{[\cdot]}$ is an equilibrium map iff for all π and any two equilibria $\sigma', \sigma'' \in \Sigma(\pi)$, $\sigma'_i \in BR_i^\pi(\sigma''_{-i})$ for all i.*

Proof: If $\sigma^{[\cdot]}$ is an equilibrium map and $\sigma', \sigma'' \in \Sigma(\pi)$, there exist sequences π_n and π_m converging to π, such that $\lim_{\pi^n \to \pi^*} \sigma_i^{[\pi^n]} = \sigma'$ and $\lim_{\pi^m \to \pi^*} \sigma_i^{[\pi^m]} = \sigma''$. Since π^n and π^m have the same limit, the definition of an equilibrium map implies that for each $\xi > 0$, there exists a $\delta > 0$ and indices N and M, such that $d(\pi^n, \pi^m) < \delta$ for all $n > N$ and $m > M$, and hence $d^s(\sigma^{[\pi^n]}, \sigma^{[\pi^m]}) < \xi$ for all $n > N$ and $m > M$. Thus, $d^s(\sigma', \sigma'') = 0$, which is equivalent to $\sigma'_i \in BR_i^\pi(\sigma''_{-i})$ and $\sigma''_i \in BR_i^\pi(\sigma'_{-i})$ for all i.

Now assume that for all π and any two equilibria $\sigma', \sigma'' \in \Sigma(\pi)$, $\sigma'_i \in BR_i^\pi(\sigma''_{-i})$ for all i. For any $\epsilon > 0$ and $\pi \in \Delta V$, let $B_\epsilon(\Sigma(\pi)) \subset \prod_i \Delta^{|A_i|}$ denote the open ϵ-ball around $\Sigma(\pi)$. By the upper hemicontinuity of Σ, $\Sigma^+[B_\epsilon(\Sigma(\pi))]$ is open for all $\pi \in \Delta V$. Hence, $\pi \in \Sigma^+[B_\epsilon(\Sigma(\pi))]$ implies that there exists an open ball around π contained in $\Sigma^+[B_\epsilon(\Sigma(\pi))]$. For each π, define

$$\delta(\pi) := \sup\{0 < \delta' < \epsilon \mid B_{\delta'}(\pi) \subset \Sigma^+[B_\epsilon(\Sigma(\pi))]\}.$$

Then the collection of sets $\left\{B_{\delta(\pi)}(\pi)\right\}_{\pi \in \Delta V}$ is an open cover of ΔV, so the Lebesgue Number Lemma[17] implies that there exists a $\delta > 0$ such that for any two points $\pi', \pi'' \in \Delta V$ with $d(\pi', \pi'') < \delta$, there exists a $B_{\delta(\pi)}(\pi)$ containing both π' and π''. Denote such a δ by $\delta(\epsilon)$.

Thus, for any $\epsilon > 0$, $d(\pi', \pi'') < \delta(\epsilon)$ implies that there exists some π such that $\pi', \pi'' \in B_{\delta(\pi)}(\pi) \subset \Sigma^+[B_\epsilon(\Sigma(\pi))]$. Therefore $\Sigma(\pi'), \Sigma(\pi'') \subset B_\epsilon(\Sigma(\pi))$, and hence $\sigma^{[\pi']}, \sigma^{[\pi'']} \in B_\epsilon(\Sigma(\pi))$. This implies that there exist $\sigma', \sigma'' \in \Sigma(\pi) \subset \Gamma(\pi)$ such that $d(\sigma^{[\pi']}, \sigma') < \epsilon$ and $d(\sigma^{[\pi'']}, \sigma'') < \epsilon$. The assumption that $\sigma'_i \in BR_i^\pi(\sigma''_{-i})$ implies that for all i, v_i and $\tilde{\sigma}_i \in \Delta^{|A_i|}$, $W_i(\pi, v_i, \sigma'_i, \sigma''_{-i}) \geq W_i(\pi, v_i, \tilde{\sigma}_i, \sigma''_{-i})$. It follows that

$$W_i(\pi, v_i, \sigma_i^{[\pi']}, \sigma''_{-i}) > W_i(\pi, v_i, \sigma'_i, \sigma''_{-i}) - \epsilon M > W_i(\pi, v_i, \tilde{\sigma}_i, \sigma''_{-i}) - \epsilon M$$

[16] See Berge (1997) for a proof.

[17] See Munkres (2000).

because $d(\sigma^{[\pi']}, \sigma') < \epsilon$;

$$W_i(\pi, v_i, \sigma_i^{[\pi']}, \sigma_{-i}^{[\pi'']}) > W_i(\pi, v_i, \tilde{\sigma}_i, \sigma_{-i}^{[\pi'']}) - 3\epsilon M$$

because $d(\sigma^{[\pi'']}, \sigma'') < \epsilon$; and

$$W_i(\pi', v_i, \sigma_i^{[\pi']}, \sigma_{-i}^{[\pi'']}) > W_i(\pi', v_i, \tilde{\sigma}_i, \sigma_{-i}^{[\pi'']}) - 5\epsilon M$$

because $\pi' \in B_{\delta(\pi)}(\pi)$ and $\delta(\pi) < \epsilon$. Letting $\xi \equiv 5\epsilon M$ for any $\xi > 0$, we get that $d^s(\sigma^{[\pi']}, \sigma^{[\pi'']}) < \xi$ whenever $d(\pi', \pi'') < \delta(\epsilon)$. ◀

Noting that the uniformity in the definition of an equilibrium map is not required for the first part of the proof, gives the following corollary:

Corollary 2 *If the sets V_i and A_i are finite for all i, then a map $\pi \mapsto \sigma^{[\pi]}$ is an equilibrium map iff for each $\pi \in \Delta V$ and each $\xi > 0$, there exists a $\delta > 0$, such that $d^s(\sigma^{[\pi]}, \sigma^{[\bar{\pi}]}) < \xi$ whenever $d^e(\pi, \bar{\pi}) < \delta$.*

The previous characterization results are very helpful to determine whether a selection from the equilibrium correspondence represents an equilibrium map. We use them in the next section, where we analyze some examples.

We conclude this section with some remarks on existence of equilibrium maps. Unfortunately, the existence question cannot be answered affirmatively for the class of finite Bayesian games analyzed in this section. As with other strong notions of equilibrium, e.g., dominant strategy or ex post equilibrium, an equilibrium map need not exist for all such games. We show this in the next section, where we present an example for which an equilibrium map does not exist.

For a given game, an affirmative answer to the existence question would support the use of models based on a common knowledge prior, with the caveat that only robust equilibria should be selected. If the answer is negative, our results indicate the necessity of a richer model, potentially based on an explicit analysis of higher order beliefs, for the analysis of such a game. Another simpler solution would be to restrict the set of feasible distributions $\text{co}_\Theta\{F(\theta)\}$ to a subset of ΔV. Whether such a restriction is possible would have to be determined on a case by case basis. For the case of finite games, it follows from the fact that the equilibrium correspondence is a semi-algebraic set (see Blume and Zame, 1994), that there always exists a non-trivial restriction for which an equilibrium map exists.

It is conceivable that existence results could be proved for more specific classes of games, e.g., games with strategic complementarities or other monotonicity properties. We leave this question for future research.

6 Examples

6.1 An Example where all Equilibrium Maps are Discontinuous

The following is a simple example for which all equilibrium maps exhibit a point of discontinuity as described in Proposition 6.

Two players must choose between two actions, a or b. The payoffs are determined by player 2's payoff type, v_2^H or v_2^L, as shown in Figure 2. It is common knowledge that player 1's payoff type is v_1 and that player 1 believes that 2's payoff type is v_2^H with probability π.

v_2^H

v_1		a	b
	a	$1,1$	$-1,1$
	b	$-1,1$	$0,0$

v_2^L

	a	b
a	$0,0$	$1,-1$
b	$1,-1$	$1,1$

Figure 2: Player 1 believes that player 2's payoff is v_2^H with probability π.

We characterize strategies using the probabilities with which players choose action a. Thus, σ_1, σ_2^H, and σ_2^L will denote the probability with which player 1, player 2 having payoff type v_2^H, and player 2 having payoff type v_2^L, respectively, choose action a.

The equilibrium correspondence as a function of π is given by the following equilibria:

1. $\sigma_1 \in \left[0, \frac{2}{3}\right)$, $\sigma_2^H = 1$, and $\sigma_2^L = 0$, for $\pi = 0$;
2. $\sigma_1 = \frac{2}{3}$, $\sigma_2^H = 1$, and $\sigma_2^L = \frac{2\pi}{1-\pi}$, for $\pi \in \left[0, \frac{1}{3}\right]$;
3. $\sigma_1 \in \left(\frac{2}{3}, 1\right)$, $\sigma_2^H = 1$, and $\sigma_2^L = 1$, for $\pi = \frac{1}{3}$;
4. $\sigma_1 = 1$, $\sigma_2^H \in \left[\frac{1}{3\pi}, 1\right]$, and $\sigma_2^L = 1$, for $\pi \in \left[\frac{1}{3}, 1\right]$.

Any selection from this equilibrium correspondence will exhibit a discontinuity in σ_1 at $\pi = \frac{1}{3}$.[18] Since at $\pi = \frac{1}{3}$, $\frac{2\pi}{1-\pi} = 1$ and $\frac{1}{3\pi} = 1$, we know that all $\sigma_1 \in \left[\frac{2}{3}, 1\right]$ are a best response to $(\sigma_2^H, \sigma_2^L) = (1,1)$, and vice versa, at $\pi = \frac{1}{3}$. It therefore follows from Proposition 6 that in order to get an equilibrium map, we can use the equilibria of class 2 and 4 above to choose a selection that is continuous on the intervals $\left[0, \frac{1}{3}\right)$ and $\left(\frac{1}{3}, 1\right]$, and pick $(\sigma_2^H, \sigma_2^L) = (1,1)$ together with any $\sigma_1 \in \left[\frac{2}{3}, 1\right]$ for $\pi = \frac{1}{3}$.

Thus, equilibrium maps exist for this game, but all such maps are discontinuous at $\pi = \frac{1}{3}$.

[18] Note also that the equilibrium correspondence fails to be lower hemicontinuous at this point.

6.2 A Modified Prisoner's Dilemma

The next example shows that not all textbook equilibria are robust. It can be interpreted as a modified Prisoner's Dilemma, where players can be either rational, with payoff type denoted by v_i^{rat}, or naive, with payoff type denoted by v_i^{naive}. Both players can either deny the accusation (action d), or confess (action c). The players' payoffs as a function of their types are illustrated in Figure 3.

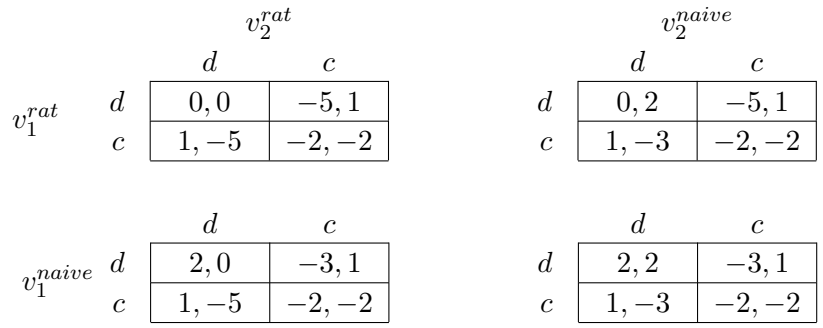

Figure 3: The true game is determined by the agents' types, $v_i \in \{v_i^{rat}, v_i^{naive}\}$.

A rational player has a dominant strategy to confess, as in the conventional Prisoner's Dilemma. A naive player gets additional utility from keeping his word and denying the accusation. Thus, if it is common knowledge that both players are naive, in addition to the Nash equilibrium where both players confess, there is an additional Nash equilibrium where both players deny.

We assume that there is incomplete information about the players' types, and denote the probability that a player i is rational by π^i. As noted before, it is a dominant strategy for a rational player to confess. Since the strategy profile where both players confess independently of their types is an ex post equilibrium, there exists an equilibrium map for this game.

We now show that even though textbook equilibria in which naive types deny with positive probability exist for some priors, such equilibria cannot lie on any equilibrium map and are therefore not robust.

Denote the probability that a naive player denies the accusation by σ_i, and denote a naive player's expected utility from denying or confessing by $u_i(d)$ and $u_i(c)$, respectively. Since rational players always confess, we have

$$u_i(d) = -3\pi^{-i} + (1-\pi^{-i})[2\sigma_{-i} - 3(1-\sigma_{-i})], \text{ and}$$
$$u_i(c) = -2\pi^{-i} + (1-\pi^{-i})[\sigma_{-i} - 2(1-\sigma_{-i})].$$

If $\sigma_{-i} = 1$, $u_i(d) \geq u_i(c)$ is equivalent to $\pi^{-i} \leq \frac{1}{2}$, so whenever π^1 and π^2 are both less then or equal to $\frac{1}{2}$, there exist pure strategy textbook equilibria with

$\sigma_1 = \sigma_2 = 1$, i.e., where both naive types deny. In addition to the pure strategy equilibria, the game also admits the following mixed strategy equilibria:

1. $\sigma_i = \frac{1}{2(1-\pi^i)}$, $\forall i$, for $(\pi^1, \pi^2) \in \left[0, \frac{1}{2}\right] \times \left[0, \frac{1}{2}\right]$;
2. $\sigma_1 = 1$, $\sigma_2 \in \left[\frac{1}{2(1-\pi^i)}, 1\right]$ for $(\pi^1, \pi^2) \in \left\{\frac{1}{2}\right\} \times \left[0, \frac{1}{2}\right]$;
3. $\sigma_1 \in \left[\frac{1}{2(1-\pi^i)}, 1\right]$, $\sigma_2 = 1$ for $(\pi^1, \pi^2) \in \left[0, \frac{1}{2}\right] \times \left\{\frac{1}{2}\right\}$.

Whenever $(\pi^1, \pi^2) \notin \left[0, \frac{1}{2}\right] \times \left[0, \frac{1}{2}\right]$, the only textbook equilibrium is such that all types confess. Since at the boundary of $\left[0, \frac{1}{2}\right] \times \left[0, \frac{1}{2}\right]$, none of the equilibrium strategies described above with $\sigma_i > 0$ constitute a best response to $\sigma_{-i} = 0$, Proposition 6 implies that no such equilibrium can be part of an equilibrium map, and therefore that such equilibria cannot be robust.

6.3 An Example where No Equilibrium Map Exists

The following is an example for which no equilibrium map with domain ΔV exists.

The payoffs are determined by the players' payoff types, $v_i \in \{v_i^H, v_i^L\}$, as shown in Figure 4. The bolded payoffs represent Nash equilibria for the individual games. As can be seen by looking at these equilibria, no ex post equilibria exist for this example.

v_2^H

v_1^H	a	b
a	**1, 1**	$-1, 1$
b	$-1, 1$	$0, 0$

v_2^L

	a	b
a	**1, 0**	$-1, -1$
b	$-1, -1$	**0, −1**

v_1^L	a	b
a	$0, 1$	**1, 1**
b	**1, 1**	$-1, 0$

	a	b
a	$0, 0$	$1, -1$
b	$1, -1$	**1, 1**

Figure 4: The true game is determined by the agents' types, $v_i \in \{v_i^H, v_i^L\}$.

We use notation introduced in previous examples, e.g., π^i denotes the probability that player i is of type v_i^H, and σ_i^H denotes the probability of choosing action a for a player of type v_i^H.

The equilibrium correspondence is given by the following sets of equilibria:

1. $\sigma_1^H = 1$, $\sigma_1^L = 1$, $\sigma_2^H \in \left[\max\left\{0, 1 - \frac{2}{3\pi^2}\right\}, 1 - \frac{1}{3\pi^2}\right]$, $\sigma_2^L = 1$, for $(\pi^1, \pi^2) \in [0,1] \times \left[\frac{1}{3}, 1\right]$;

2. $\sigma_1^H = 1$, $\sigma_1^L = 0$, $\sigma_2^H = 1$, and

$$\sigma_2^L \begin{cases} = 0, & \text{for } (\pi^1, \pi^2) \in \left[0, \frac{2}{3}\right) \times \left[\frac{1}{3}, 1\right] \\ \in \left[\frac{1-3\pi^2}{3(1-\pi^2)}, 1\right] & \text{for } (\pi^1, \pi^2) \in \left\{\frac{2}{3}\right\} \times [0,1] \\ = 1, & \text{for } (\pi^1, \pi^2) \in \left(\frac{2}{3}, 1\right] \times [0,1] \end{cases};$$

3. $\sigma_1^H = 0$, $\sigma_1^L = 0$, $\sigma_2^H = 1$, $\sigma_2^L = 0$, for $(\pi^1, \pi^2) \in [0,1] \times \left[0, \frac{1}{3}\right]$;

4. $\sigma_1^H = \frac{2(1-\pi^1)}{\pi^1}$,$\sigma_1^L = 0$, $\sigma_2^H = 1$, $\sigma_2^L = \frac{1-3\pi^2}{3(1-\pi^2)}$, for $(\pi^1, \pi^2) \in \left[\frac{2}{3}, 1\right] \times \left[0, \frac{1}{3}\right]$.

A rather tedious comparison of the previous classes of equilibria using Proposition 6, shows that no equilibrium map exists for this example.

6.4 Cournot Duopoly with Incomplete Information

Our final example is a simple Cournot duopoly game with incomplete information. It shows that non-constant equilibrium maps, defined by pure strategies only, can exist when the set of available actions is a continuum.

Two firms, $i \in \{1, 2\}$, compete in quantities, q_i, in a market with inverse demand given by $p(q) = a - q$, where $q = q_1 + q_2$. There is incomplete information about the firms' marginal cost of production, which can be either high, c_i^H, or low, c_i^L. We denote the probability that firm i's marginal cost is high by π^i.

Letting $q_i(c_i)$ denote firm i's supply as a function of its cost, we get the following first-order conditions for any firm i's profit maximization problem:

$$q_i(c_i^H) = \frac{1}{2}\left\{\pi^{-i}[a - q_{-i}(c_{-i}^H) - c_i^H] + (1 - \pi^{-i})[a - q_{-i}(c_{-i}^L) - c_i^H]\right\},$$
$$q_i(c_i^L) = \frac{1}{2}\left\{\pi^{-i}[a - q_{-i}(c_{-i}^H) - c_i^L] + (1 - \pi^{-i})[a - q_{-i}(c_{-i}^L) - c_i^L]\right\}.$$

It is easy to see that if assumptions are made to guarantee a unique equilibrium for each (π^1, π^2), the firms' equilibrium supplies will be continuous in (π^1, π^2). In this case, the equilibrium correspondence is continuous, not constant, and it defines an equilibrium map for this game.

7 Concluding Remarks

7.1 Relation to the Literature

In the paper that introduced common p-belief, Monderer and Samet (1989) show that the assumption that a given game is common knowledge can be successfully

replaced with the assumption that the game is common p-belief, thereby showing that common p-belief is an appropriate notion of approximate common knowledge. Using a state space model in which the true state of nature determines which one of a finite number of finite, normal form, complete information games is played, they show that Nash equilibrium strategies for any game $\mathcal{G}$ can constitute an ε-best response at states of nature at which it is common p-belief that game $\mathcal{G}$ is played.

We show that common p-belief can also be used to relax the assumption of a common knowledge prior for Bayesian games. Unlike Monderer and Samet (1989), where it is assumed that the exact game is common p-belief, we only assume that it is common p-belief that players derive their beliefs from priors which are not too different. This is important, since in our framework, the assumption used in Monderer and Samet (1989) would imply that with high probability, all players have exactly the same prior. Given that the set of priors is an infinite set, this would not yield a realistic relaxation of the common prior assumption. The cost of weakening this assumption is that we need to introduce the concept of an equilibrium map in order to define an equilibrium.

As noted in the discussion of Example 1, the equilibrium correspondence that maps common knowledge priors to the corresponding set of textbook equilibria, need not be lower hemicontinuous. Kajii and Morris (1998) define a distance on the set of common knowledge priors under which the ε-equilibrium correspondence is lower hemicontinuous. Thus, their result shows that if all players coordinate on an a textbook equilibrium derived for some common knowledge prior, such an equilibrium is an ε-equilibrium if types are distributed according to a prior that is close according to this distance. Note that players have to coordinate on a textbook equilibrium for the same prior. Therefore, the result of Kajii and Morris (1998) applies to the case where all players' beliefs are derived from the same prior, which may be incorrect, whereas our paper models the case where players' beliefs may be derived from different priors.

Since in the model of Kajii and Morris (1998), both utilities and strategies are a function of the types over which the priors are defined, their distance can be interpreted as providing a notion of closeness for information structures corresponding to different common knowledge priors on payoff types. In contrast, the perturbation we define models closeness of arbitrary belief structures to common knowledge prior beliefs. Although the distance defined in Kajii and Morris (1998) and the distance inherent in our definition of a perturbation are both based on the notion of common p-belief, a direct comparison is problematic due to differences in the modelling framework.

One significant difference is that Kajii and Morris (1998) do not allow for be-

liefs that are not derived from a common knowledge prior, so the players in their model can never believe that there is a possibility that their opponents beliefs are not consistent with the given prior. We relax this assumption by introducing belief types to model both differences in priors on payoff types, and beliefs about such differences. By considering payoff types and belief types separately, we can distinguish between textbook games and games based on more complicated type spaces, which is not possible in the framework of Kajii and Morris (1998).

A different notion of robustness, which is also based on properties of the equilibrium correspondence, is that of an `essential equilibrium` (see Fudenberg and Tirole, 1991, pp. 480-484). In our framework, a textbook equilibrium is essential if for any common knowledge prior close to the one used to derive the equilibrium, there exists a corresponding textbook equilibrium that is close to the original equilibrium, where closeness is measured using the standard topologies on strategies. It follows from the fact that the equilibrium correspondence is continuous on a residual of ΔV, that for all priors in this residuals, all corresponding textbook equilibria are essential.

The main conceptual difference between essential equilibria and our definition of robustness is that higher order beliefs are not taken into account when defining essential equilibria. Thus, essential equilibria can be interpreted as being robust to a perturbation of the common prior which is identical across players, and which maintains that the perturbed prior is common knowledge.

Note also that the continuity requirement in the definition of an essential equilibrium is stronger then the strategic continuity used in our definition of an equilibrium map. Hence, a textbook equilibrium could be robust according to our definition without being an essential equilibrium.

7.2 Conclusion

This paper introduced a notion of robust equilibria for Bayesian games. We looked at a perturbation of those beliefs defined by a common knowledge prior, a perturbation that includes a large set of higher order beliefs. For beliefs in this perturbed set, we showed how robust equilibria can be defined by a selection from the equilibrium correspondence which is a function of priors on payoff types. Since such a selection is independent of higher order beliefs, it can be characterized without resorting to the use of complicated type spaces.

Our results can be applied as part of any analysis involving Bayesian games, e.g., auctions, contracts, or any general mechanism design problem. In contrast to the classical mechanism design literature, where optimal mechanisms can be very dependent on the assumed common knowledge prior, our notion of robustness

could potentially be used to derive mechanisms that are approximately optimal for a variety of beliefs.

Appendix

Proof of Proposition 1: (ii) We can write $E = E_i \times T_{-i}$, for some $E_i \in T_i$. Property 4 of the definition of a type space implies that $\mu_i(t_i)(E) = 1$ if $t_i \in E_i$, and 0 otherwise, which implies (ii).

(iii) If $E \subseteq F$, then $\mu_i(t_i)(E) \geq p$ implies that $\mu_i(t_i)(F) \geq p$, so $B_i^p(E) \subseteq B_i^p(F)$.

(iv) This follows from (iii).

(v) $B_i^p(E \cap F) \subseteq E \cap B_i^p(F)$ is an easy consequence of (iii) and (ii). For the other direction, note that $E \in \mathscr{F}_i$ and $(t_i, t_{-i}) \in E \cap B_i^p(F) = B_i^p(E) \cap B_i^p(F)$ implies $\mu_i(t_i)(E) = 1$ and $\mu_i(t_i)(F) \geq p$. Hence, $\mu_i(t_i)(F \setminus E) = 0$ and so $\mu_i(t_i)(E \cap F) \geq p$. ◄

Proof of Lemma 1:

$$\begin{aligned} B^p(B^p(E)) &= \bigcap_i B_i^p(B^p(E)) = \bigcap_i B_i^p \left[\bigcap_i B_i^p(E) \right] \\ &= \bigcap_i \left\{ B_i^p(E) \cap B_i^p \left[\bigcap_{j \neq i} B_j^p(E) \right] \right\} \\ &\subseteq \bigcap_i B_i^p(E) = B^p(E), \end{aligned}$$

where the third equality follows from Proposition 1, (i), (ii) and (v). ◄

Proof of Lemma 2: $\forall n \geq 1$, $\mathcal{C}^p(E) \subseteq [B^p]^{n+1}(E) = \bigcap_i B_i^p([B^p]^n(E)) \subseteq B_i^p([B^p]^n(E))$. Hence $\mathcal{C}^p(E) \subseteq \bigcap_{n \geq 1} B_i^p([B^p]^n(E))$ and the previous Lemma together with (iv) of Proposition 1 imply that $\forall i$, $\mathcal{C}^p(E) \subseteq B_i^p(\bigcap_{n \geq 1} [B^p]^n(E)) = B_i^p(\mathcal{C}^p(E))$, so $\mathcal{C}^p(E) \subseteq B^p(\mathcal{C}^p(E))$. ◄

Proof of Proposition 2: ($\Leftarrow$) $\mathcal{C}^p(E)$ is an evident p-belief event by Lemma 2, and since $\mathcal{C}^p(E) \subseteq B^p(E)$ by definition, it follows that $E \in \mathscr{F}$ is common p-belief at every $t \in \mathcal{C}^p(E)$.

($\Rightarrow$) By assumption, there exists an $F \in \mathscr{F}$ such that $t \in F$, $F \subseteq B^p(F)$ and $F \subseteq B^p(E)$. We show by induction that $F \subseteq [B^p]^n(E)$ for all $n \geq 1$: This holds by assumption for $n = 1$, so assume that $F \subseteq [B^p]^n(E)$ for some $n \geq 1$; Property (iii) from Proposition 1 implies that $B^p(F) \subseteq [B^p]^{n+1}(E)$, so $F \subseteq [B^p]^{n+1}(E)$. Since $t \in F$, it follows that $t \in \mathcal{C}^p(E)$ ◄

References

Aubin, J.-P., and H. Frankowska (1990): *Set-Valued Analysis*. Birkhäuser.

Berge, C. (1997): *Topological Spaces*. Dover.

Bergemann, D., and S. Morris (2003): "Robust Mechanism Design," Cowles Foundation Discussion Paper No. 1421, Yale University.

Blume, L. E., and W. R. Zame (1994): "The Algebraic Geometry of Perfect and Sequential Equilibrium," *Econometrica*, 62(4), 783–794.

Chung, K.-S., and J. C. Ely (2002): "Ex-Post Incentive Compatible Mechanism Design," CMS-EMS Discussion Paper No. 1339, Northwestern University.

Engl, G. (1995): "Lower Hemicontinuity of the Nash Equilibrium Correspondence," *Games and Economic Behavior*, 9(2), 151–160.

Folland, G. B. (1999): *Real Analysis: Modern Techniques and Their Applications, Second Edition*. Wiley-Interscience.

Fudenberg, D., and J. Tirole (1991): *Game Theory*. MIT Press.

Harsanyi, J. C. (1967-68): "Games with Incomplete Information Played by Bayesian Players, I-III," *Management Science*, 14, 159–182, 320–334, 486–502.

Heifetz, A., and D. Samet (1998): "Topology-Free Typology of Beliefs," *Journal of Economic Theory*, 82(2), 324–341.

Kajii, A., and S. Morris (1997): "Common p-Belief: The General Case," *Games and Economic Behavior*, 18(1), 73–82.

——— (1998): "Payoff Continuity in Incomplete Information Games," *Journal of Economic Theory*, 82(1), 267–276.

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

Milgrom, P. R., and R. J. Weber (1985): "Distributional Strategies for Games with Incomplete Information," *Mathematics of Operations Research*, 10(4), 619–632.

Monderer, D., and D. Samet (1989): "Approximating Common Knowledge with Common Beliefs," *Games and Economic Behavior*, 1(2), 170–190.

Munkres, J. R. (2000): *Topology, Second Edition*. Prentice Hall.

Rubinstein, A. (1989): "The Electronic Mail Game: Strategic Behavior Under "Almost Common Knowledge"," *American Economic Review*, 79(3), 385–391.

Stroock, D. W. (1999): *Probability Theory, An Analytic View, Revised Edition*. Cambridge University Press.

Wilson, R. B. (1987): "Game-Theoretic Analyses of Trading Processes," in *Advances in Economic Theory: Fifth World Congress*, ed. by T. Bewley, chap. 2, pp. 33–70. Cambridge University Press.